THE UNIVERSITY *of* EDINBURGH

An improved description of Earth's external magnetic fields and their source regions using satellite data

Robert Michael Shore

Thesis submitted for the Degree of

Doctor of Philosophy

The University of Edinburgh

2013

I declare that this thesis has been composed solely by myself, and that it contains only my work except where otherwise specified, or where the work is explicitly indicated below to have formed part of a jointly-authored publication. This work has not been submitted for any other degree or professional qualification.

Chapter 3 is based on jointly-authored work also published as Shore et al. (2013). The work in chapter 3 and Shore et al. (2013) was done by me. The contributions of the co-authors are as follows. My supervisors Kathy Whaler, Susan Macmillan and Ciaran Beggan provided feedback on my descriptions of ionospheric theory and interpretation of the currents, guidance on how to develop the research to completion, and multiple iterations of proofreading the publication manuscript. Anasuya Aruliah provided ionospheric theory advice. Tim Spain provided data predictions from the CTIP model. Nils Olsen provided assistance on the theory of modelling satellite attitude errors and how these might be used to improve the conclusions of the work, as well as synthetic *Swarm* constellation data and code for various coordinate system rotations.

I confirm that appropriate credit has been given within this thesis where reference has been made to the work of others.

Robert Michael Shore

Abstract

In near-Earth space, highly spatio-temporally variant magnetic fields result from solar-terrestrial magnetic interaction. These near-Earth external fields currently represent the largest source of error in efforts to model the magnetic field produced in the Earth's interior. Starting in 1999, the Decade of Geopotential Field Research (Friis-Christensen et al., 2009) has greatly increased the amount of available low-Earth orbit (LEO) satellite magnetic data. These data have driven many advances in field modelling, yet have highlighted that LEO measurements are particularly susceptible to contamination from external fields. This thesis presents a series of studies attempting to describe the external fields in more detail, in order that they can be more effectively separated from the internal fields in magnetic modelling efforts. A range of analysis methods, different for each study, are applied to satellite and ground-based observatory data.

Mandea and Olsen's (2006) method of estimating the secular variation (SV) of the internal field from satellite data via 'Virtual Observatories' (VOs) is applied to synthetic data from the upcoming *Swarm* constellation satellite mission of the European Space Agency. Beggan (2009) found VOs constructed from CHAMP satellite data to be contaminated with external field signals which appeared to have a significant local time (LT) dependence. I find that utilising the increased coverage of LT sectors offered by the *Swarm* constellation geometry does not significantly decrease the contamination. Following this surprising result I tested a wide range of methods aimed at reducing the VO contamination from each parameterised external field source region. In anticipation of future studies using real data, I used the results of the tests to provide a more complete description of the external field variations affecting analyses of geographically-fixed magnetic phenomena when using satellite data and spherical harmonic analysis (SHA).

Ionospheric electric currents flowing at LEO altitudes are known to violate the assumption of measurements taken in a source-free space, required in SHA-based models of the magnetic field. In order to better describe the electromagnetic environment at LEO altitudes, I use data from the Ørsted and CHAMP satellites to calculate the current density from Ampère's integral.

Vector magnetic data from discrete overflights of the two satellites (at different altitudes) are rotated into the along-track frame to define the integral loop and its 'surface area', permitting estimation of the predominantly zonal current density flowing in the region between the two orbital paths. I designed selection criteria to extract geometrically-stable overflights spanning the range of LTs twice in the 6 years of mutually available satellite vector data. From these overflights I resolve current densities in the range $\pm 0.1\ \mu\mathrm{A/m^2}$, with the distribution of current largely matching the LT progression of the Appleton anomaly. I applied detailed tests to check for biases intrinsic to the method, and present results free of systematic errors. The results are compared with the predictions of the CTIP (Coupled Thermosphere-Ionosphere-Plasmasphere) model of ionospheric composition and temperature, showing a typically good spatiotemporal agreement. I find persistent current intensifications between geomagnetic latitudes of 30° and 50° in the post-midnight, pre-dawn sector, a region which has been previously considered to be relatively free of currents.

External fields induce currents in the Earth's conducting mantle, the magnetic fields of which add to the field measured at and above the Earth's surface. The morphology of the long-period inducing field is poorly resolved on timescales of months to years, reducing the accuracy of mantle induction studies (a key part of the *Swarm* mission). I improve the description of its morphology via the method of Empirical Orthogonal Functions (EOFs), which I apply to over a decade of ground-based observatory data. EOFs provide a decomposition of the spatio-temporal structures contained in the magnetic field data, with partitions arising from the data themselves, overcoming the relatively simplistic assumptions made about the inducing field morphology in LT. The results of vector data EOF analyses are presented, but I rely primarily on scalar analyses which are more fitting for this study. I overcome the limitations of the irregular observatory distribution with a novel spatial weighting matrix, combining the output from multiple EOF analyses to greatly improve the data coverage in LT. I find that the seasonal variation of the inducing field is more important than the variation of the symmetric ring current on annual periods, and that dawn-dusk asymmetry should be accounted for to increase the accuracy of mantle conductivity estimates based on data covering the decadal timescales of the solar cycle.

Lay summary

The most substantial part of the Earths magnetic field – that makes compasses point approximately North – is produced by motions of the Earth's liquid iron outer core, deep in its interior, so studying the behaviour of this field provides insights into processes occurring there. Models (mathematical representations) of this internal field have wide ranging applications from commercial navigation to resource exploration. However, one of the greatest challenges in modelling this field is distinguishing between it and other magnetic fields generated above the Earth's surface in the upper atmosphere and in space – these are known as external fields. External fields are weaker, but change more quickly, both in time and space, than internal fields. Starting in 1999, several satellite missions were launched equipped with high-precision magnetic field measuring instruments. These have greatly increased the amount of magnetic data available from low-Earth orbit (LEO – between about 250 and 800 km altitude). These data have driven many scientific advances, yet have highlighted that LEO measurements are particularly affected by external fields. My thesis presents a series of studies attempting to describe the external fields in more detail, in order that they can be more effectively separated from the internal fields.

Satellite data supplement the permanently-running network of ground-based magnetic observatories, and have proven very useful in measuring how the internal field changes with time in areas where data were not previously available (*e.g.* oceans). In 2006, Mandea and Olsen developed a method of mimicking the measurements of ground-based observatories with satellite data to construct a 'Virtual Observatory' (VO) at an altitude of 400 km, providing near-global coverage. However, there were some problems with this method which my thesis investigates using synthetic data from the upcoming Swarm mission (of the European Space Agency). The Swarm mission will have three satellites orbiting simultaneously, two in low orbit (450 km) and one at a higher orbit (530 km). I use synthetic data (predictions of the satellite measurements) for Swarm to provide detailed quantification of how external fields affect the VO value. This new information will allow greater accuracy in future studies of internal fields with real satellite data.

Another obstacle in the use of satellite data for studying magnetic fields is the presence

of electrical currents in the Earths upper atmosphere, between 100 and 1000 km altitude. The upper atmosphere can conduct electricity and the resulting electrical currents generate magnetic fields, affecting LEO magnetic measurements. I have helped improve the processing of satellite data by customising and implementing a method of estimating this electric current. To do this, I use measurements from two previous satellite missions (called Ørsted and CHAMP, respectively) where they simultaneously measured the magnetic field (though at different altitudes).

On timescales of months and greater, the external fields generate secondary magnetic fields in the regions of the Earth's interior outside of the iron core (via a process called induction). These smaller secondary fields add to the fields measured at and above the Earth's surface. However, the geometry of the external field which causes this induction is not clearly defined. I analysed over a decade of ground-based observatory data using a method called Empirical Orthogonal Functions (EOFs), designed to isolate patterns in space and time which are contained within the observatory records. EOFs allow the information contained in the magnetic field to provide a representative breakdown of the observatory measurements, rather than scientists assuming a particular pattern and then estimating its strength. This improved geometry of the external field on long timescales will lead to more accurate studies of the Earth's interior which utilise the information from induction processes.

Acknowledgements

First and foremost, I would like to thank my supervisors for all their help throughout my Ph.D. Their critiques of the many presentations of research-in-progress and write-up drafts helped me stay on track as the project developed. They have each provided me invaluable guidance and encouragement, and were a large part of making this project a success. In particular I would like to thank my primary supervisor Kathy Whaler, who bore the brunt of me heading to her office unannounced for discussions of research results and direction. Kathy was supportive throughout the entirety of my Ph.D. and never made me feel like the project was something I could not achieve. I feel lucky to have had her as a supervisor. I thank my parents for getting me to the stage at which I could undertake this project and for their encouragement, and that of my family, throughout the process. Thanks go to Jennifer Wright for her advice and encouragement during the writing-up stage, and for generally making my time at Edinburgh a better experience.

Computer coding assistance was gratefully received from Ciaran Beggan, Matthew Unterman, Louise Barron, Ruth Carley, Roger Hipkin, Simon King, Nick Johnson, Nil Olsen, Dieter Werthmüller and Antony Bloom. I would like to thank the following people for theoretical assistance with mathematical and geophysical concepts: my supervisors, Nils Olsen, Hermann Lühr, Simon Tett, Roger Hipkin, Andrew Curtis, Rod Heelis, Anasuya Aruliah, Tim Fuller-Rowell and Ruth Carley. I used data from a multitude of sources throughout the project and would like to thank the CHAMP and Ørsted data centres for provision of satellite magnetic data, Tim Spain for providing CTIP model data used in chapter 3, Olsen *et al.* for their work on (and provision of) the E2Eplus *Swarm* data (Olsen et al., 2007), and Olsen *et al.* for the provision of the ground-based observatory data described in the ESA *Swarm* Level 2 processing report (Olsen et al., 2011), used in chapter 4.

Lastly, I would like to thank the inhabitants of the Grant Institute and Crew Building postgraduate student attics, and all the staff and students who came down to coffee in both buildings, for making Edinburgh University a very pleasant place in which to work.

Contents

List of Figures

List of Tables

List of Equations

Chapter 1

Background

1.1 Glossary

Table 1.1: Glossary of terms used in this chapter

Abbreviation	Explanation
R_E	Earth radius
IMF	Interplanetary magnetic field
FAC	Field-aligned current
LEO	Low-Earth orbit
SV	Secular Variation (temporal change of the core field, typically on scales of decades and centuries but here on the scale of month and years)
LT	Local time
yr	Years
a	Years ago (*e.g.* Ma, Ga for million, billion)

1.2 Basics of terrestrial magnetism

1.2.1 The Solar-terrestrial environment

The Earth is a differentially-layered rocky planet with a mean radius of 6371 km, in orbit around the Sun at a distance of roughly 1.5×10^8 km. The planet accreted 4.5 Ga (billion years ago) from smaller planetisimals in the solar system's protoplanetary disc in a process lasting roughly 30 Myr (million years). Late in this formation process, the proto-Earth collided with a Mars-sized protoplanet named Theia, providing enough heat energy to completely remelt the Earth, and sufficient silicate ejecta to form the Earth's relatively iron-poor Moon (Wood et al., 2006).

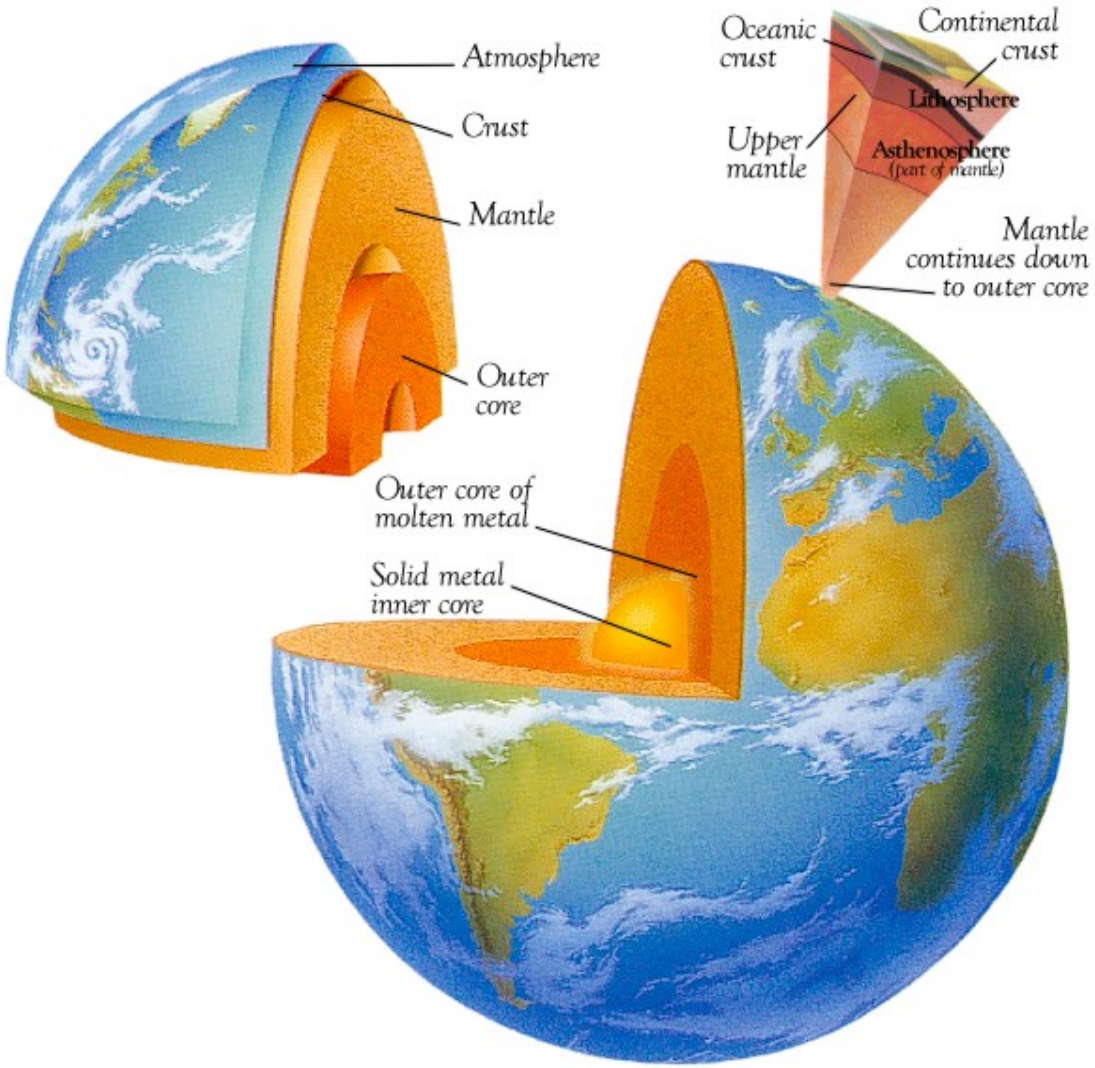

Figure 1.1: Schematic illustration of the layers of the Earth. More detail is given in text. From `http://library.thinkquest.org/C003124/en/fullstruct.htm`, accessed 11/05/2013.

The layering of the Earth is shown schematically in Figure 1.1. Starting from the surface and progressing through each (simplified) differentiated layer towards the centre of the Earth, we first encounter the crust, a solid layer with a thickness between 5 and 30 km, being thinner under the oceans and thicker under continental regions (Campbell, 2003, page 52). The crust is the uppermost part of the lithosphere, a relatively rigid layer extending to a depth of 100 km, and at the top of the Earth's silicate mantle (Langel & Hinze, 1998, page 2). Beneath the litho-

sphere, the upper mantle is termed the asthenosphere, being relatively more plastic than the layers above it. At a depth of 600-700 km, the upper mantle transitions to the denser lower mantle which extends to a mean depth of 2890 km, that of the core-mantle boundary (CMB). The (predominantly iron) core is divided into two parts: the liquid outer core extends to a depth of 5150 km, and the solid inner core occupies the remaining centre of the Earth.

The Earth has exhibited a self-sustaining magnetic field for much of its 4.5 Gyr history, with the earliest recorded field currently at 3.2 Ga (Tarduno et al., 2007). This magnetic field is generated from the convective motion of the liquid iron in the outer core, driven mainly by the release of buoyant 'impurities' (*e.g.* O, S) from the continued freezing out of the inner core (Glatzmaier & Olson, 2005). However, owing to the relatively young age of the inner core (suggested by Buffett (2003) to have formed approximately 1 Ga), thermal convection must have driven core flow early in the Earth's history and will also contribute to magnetic field generation now. In a source-free region, the full vector magnetic field can be decomposed such that it is defined by two independent vector fields (*e.g.* Backus, 1986; Jones et al., 2010), called the poloidal and toroidal parts of the field. The poloidal component is the radial part of the magnetic field, whilst by definition the toroidal field has no radial component. As discussed by Jones et al. (2010); Jones (2011), azimuthal motion of the electrically-conducting liquid iron in an existing poloidal magnetic field (presumed to be initially of extraterrestrial origin) creates strong electric currents, with associated toroidal magnetic fields. If the mantle were perfectly insulating (which it is not), the toroidal field would not leave the core – in practice it is not distinguishable at or above the Earth's surface. Columnar convection (driven by Earth rotation) of the outer core fluid causes lines of toroidal field to generate a renewed poloidal field (Olson et al., 1999) – it is this radial magnetic field which permeates the mantle and is measurable at and above the Earth's surface. Here it has a strength of 60,000 nT near the magnetic poles, and 30,000 nT near the dip equator (at which the radial component of the field has zero magnitude) (Campbell, 2003, page 7). The temporal change of the core magnetic field is known as the secular variation (SV). At present the SV of the dipole component of the poloidal field is -11 nT/yr (Finlay et al., 2010).

The conversion of toroidal to poloidal magnetic field (and vice versa) also accounts for the generation of the Sun's poloidal magnetic field (Cattaneo & Hughes, 2001). The poloidal magnetic fields from both the Sun and the Earth extend outwards into their near-space environments. However, the Sun's magnetic field (called the interplanetary magnetic field, or IMF) is 'carried' in a supersonic stream of plasma that is emitted from the Sun, termed the solar wind. The field-line geometry of the IMF is shown in Figure 1.2. As the solar wind plasma meets the region of influence of the Earth's magnetic field it slows to subsonic speeds, creating a continuous, collision-free shockwave, named the bow shock. Slightly inside this shock is the boundary at which the solar wind dynamic pressure and terrestrial magnetic field forces balance, called the magnetopause – it has a mean distance of 11-12 R_E (Earth radii) from the Earth (Campbell, 2003). At this boundary the solar magnetic field (interplanetary magnetic field or IMF) has a magnitude of about 5 nT, and the terrestrial magnetic field a magnitude of around 20 nT (Backus et al., 1996)[1]. The region contained within the magnetopause is a cavity in the flow of the solar wind, formed by the sunwards/anti-sunwards compression/extension of the Earth's dipole-dominated poloidal field lines into a structure reminiscent of the particle-stream of a comet (shown schematically in Figure 1.3). This region is termed the magnetosphere, and it is host to several dynamic collections of charged particles with associated electric current systems. These collections of charged particles arise either from influx of solar wind particles into the magnetosphere or from upflow of ionised plasma out of the Earth's atmosphere along the magnetic field lines. The latter source of charged particles results from the irradiation-related ionisation of in-situ neutral particles in the Earth's upper atmosphere: this region from 90–1000 km above the Earth's surface is called the Earth's ionosphere (Kelley, 2009). The magnetic fields caused by charged particle motions in both the ionosphere and magnetosphere produce measurable (occasionally strong) disturbances in the magnetic fields recorded on the Earth's surface, and above it by satellites in low-Earth orbit (LEO). The magnetospheric and ionospheric magnetic fields are referred to collectively as the Earth's 'external fields'. Due to their variation with time, the external fields have induction effects in the conducting mantle,

[1] terrestrial magnetic field computed using the International Geomagnetic Reference Field model (Finlay et al., 2010) at 11 R_E, at the geographic equator and 0 degrees geographic longitude

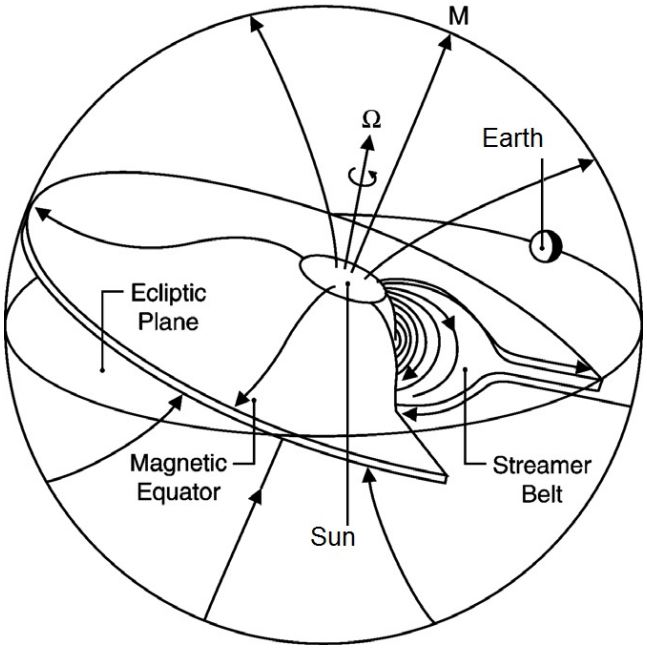

Figure 1.2: Schematic illustration (not to scale) of the IMF generation from the Sun's poloidal magnetic field lines. Note the neutral current sheet separating the two hemispheres of magnetic polarity. As the IMF impinges on the Earth, its apparent direction changes according to the position of the current sheet. The tilt of the solar magnetic dipole (M) with respect to the Sun's rotation axis (Ω) causes an apparent polarity change in the IMF with a 27-day period (according to solar rotation) at the Earth's orbital distance. After Russell & Jian (2008).

lithosphere and oceans. The resulting magnetic fields add to the internal field measured at and above the Earth's surface. I discuss lithospheric (crustal) fields in chapter 2, and mantle fields are discussed in chapter 4, whilst oceanic magnetic fields are small enough to be considered negligible.

The vector property of the Earth's magnetic field has motivated its use in global navigation for several hundred years (Backus et al., 1996), and magnetic field navigation is important even today in the oil and gas exploitation, aviation and shipping industries. The accurate study of the core field is of scientific interest in its own right, allowing us an insight into processes occurring deep within the Earth. In turn, a better knowledge of the deep-Earth processes which create the internal field allows for further improvements to the accuracy of surface applications such as navigation, and also improves the forecast potential of internal field models in future epochs. In this thesis I study the external fields both at and above the Earth's surface, aiming to

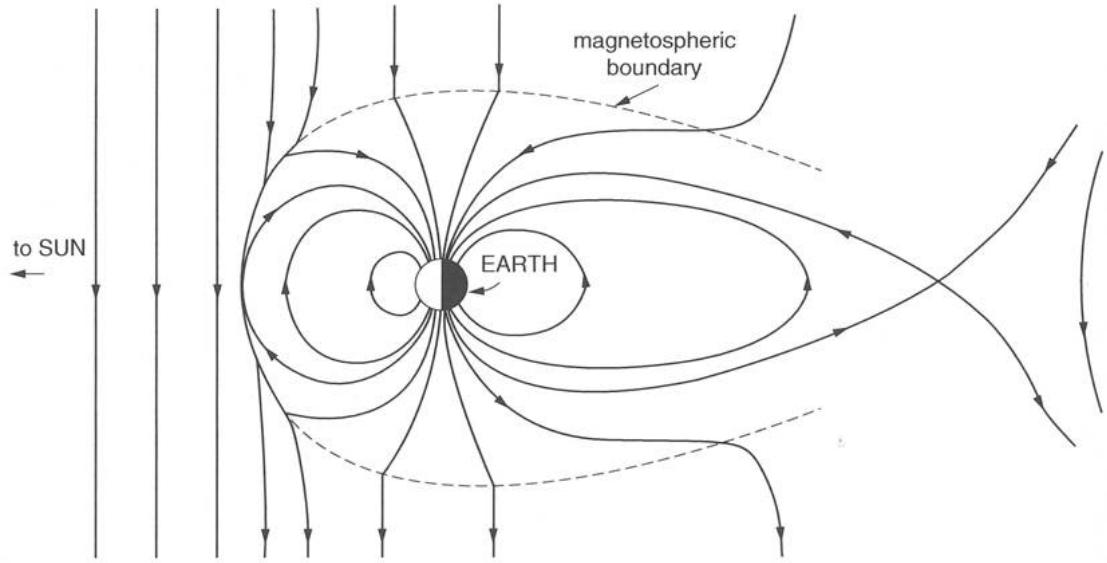

Figure 1.3: Schematic illustration of the compression and extension of Earth's dipole-dominated magnetic field lines by the (here, southwards) IMF, forming the magnetosphere. From Campbell (2003).

improve the description of the external fields in order to better mitigate their negative impacts on efforts to characterise the magnetic field internal to the Earth. This is a rather broad aim as the interaction of the terrestrial and solar magnetic fields spans a wide variety of spatial and temporal scales, to say nothing of the complex scope implied in the modelling of the Earth's internal field. Before discussing the more precise aims of this research, I will supply some essential background detail.

1.2.2 Spatial variations of the Earth's external magnetic fields

In this sub-section I describe the approximate magnitudes of the external magnetic fields (at the Earth's surface during quiet times unless otherwise stated), and their arrangement in local time (LT) and latitude in order to describe their apparent diurnal variation resulting from Earth rotation. I study the external fields both at and above the Earth's surface, but in most cases I am able to treat the magnetic fields generated by the current systems as being far from the source (i.e largely independent of the internal structure of the source). Thus here I will focus on the bulk geometry of the separate current systems, and will not discuss differences in the magnetic fields within these regions unless it affects their signal in my data sources. The exception to

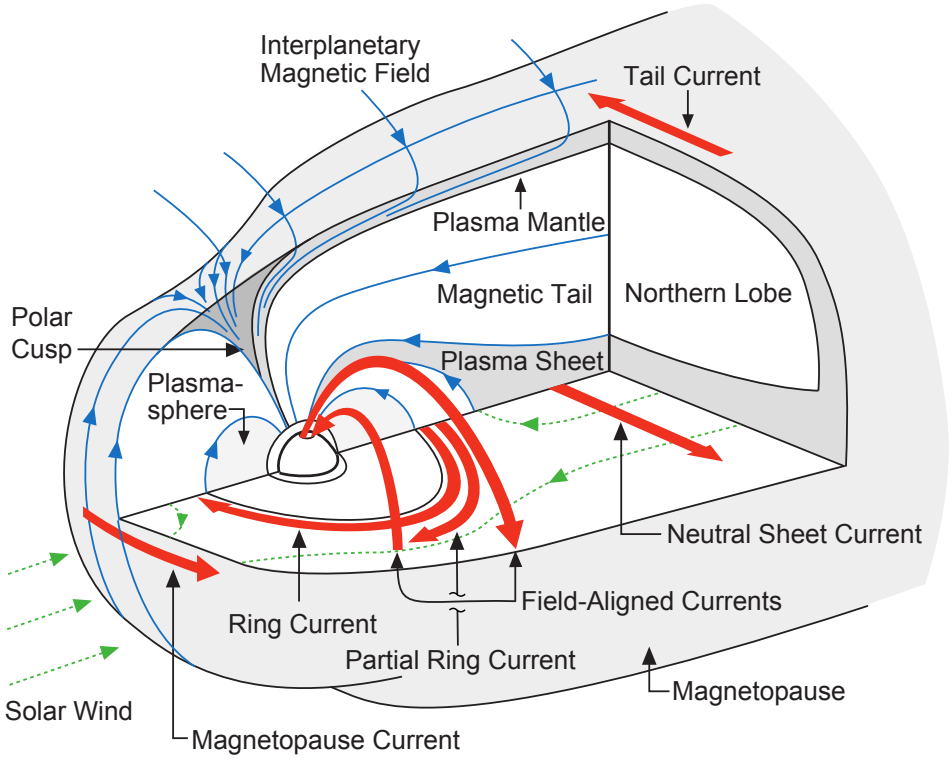

Figure 1.4: Cut-away schematic of the magnetosphere. Blue lines indicate magnetic field lines, green lines indicate plasma flow, shading indicates collections or cavities of charged particles. Red arrows indicate major magnetospheric current systems responsible for the structure of the external magnetic fields in local time. Image modified from Russell (2000).

this is my study of in-situ ionospheric electric currents (chapter 3), and I discuss the internal structure of the ionosphere in more detail in that chapter.

A simplified schematic of the internal magnetospheric structure is shown in Figure 1.4. The major magnetospheric current system, and one thought until recently to define almost the entirety of the magnetosphere, is the symmetric ring current (called symmetric because it is largely invariant in LT, and labelled 'Ring Current' in Figure 1.4), which encircles the Earth in the magnetic-equatorial plane between 3–5 R_E (Kivelson & Russell, 1995, page 407). The ring current is formed of plasma particles which gyrate around the Earth's magnetic field lines. Due to the convergence of the field lines with latitude, the gyrating particles are trapped, 'bouncing' between the conjugate ends of magnetospheric flux tubes as shown in Figure 1.5. The bounce

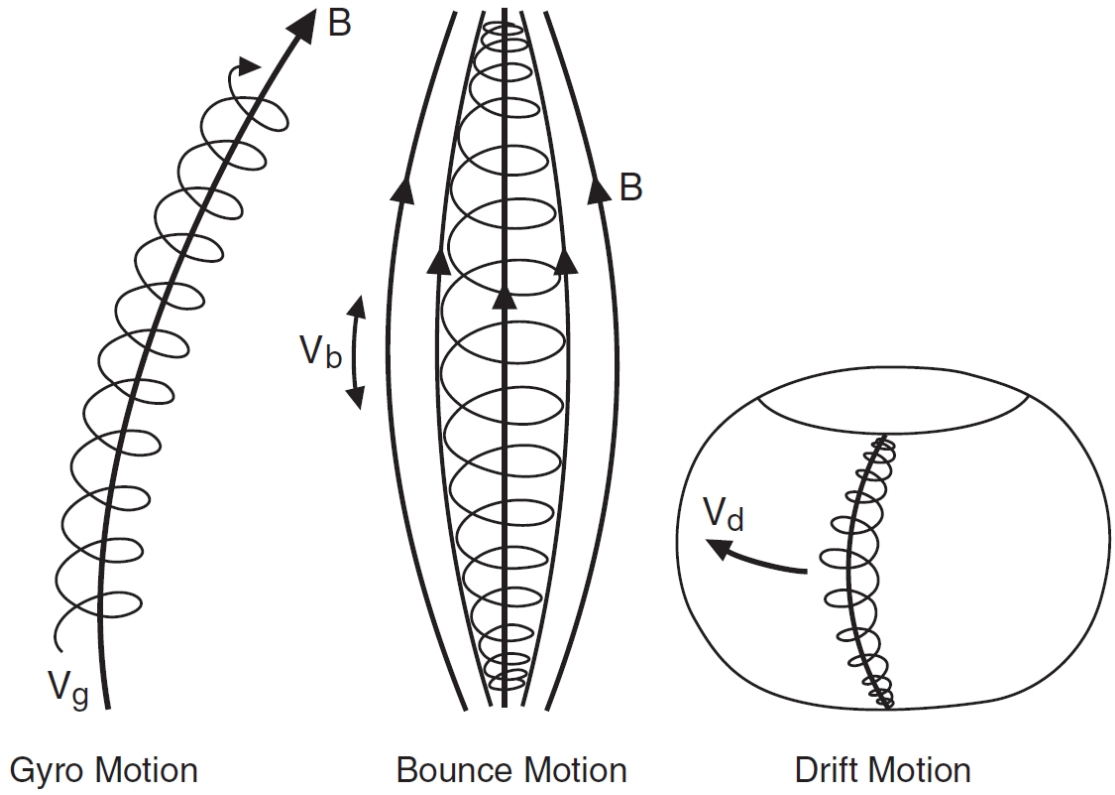

Figure 1.5: Schematic illustration of the motion of charged particles around magnetic field lines (**B**) which contribute to the magnetospheric ring current (shown as a torus in the right-hand image). The left-hand image shows the gyration motion (with gyration velocity v_g) of a charged particle around a magnetic field line due to the Lorentz force (Kivelson & Russell, 1995). If the field lines converge (as in the centre image), the increasing magnetic field strength causes a force opposite to the direction of the particle's motion along the field line. The particle will then 'bounce' back and forth along the field line with bounce velocity $\pm v_b$. The bounce and drift velocities of the charged particles trap them in radiation belts with a toroidal geometry. Both the curvature of the magnetic field lines, and the decrease in field strength with distance from the Earth, cause the trapped particles to drift with velocity v_d. This drift motion is westwards for ions (as shown in the left-hand image) and eastwards for electrons. Thus a westwards ring current is created which encircles the Earth. Image and details from Russell (2000).

motion is rapid, typically of period 10 seconds for electrons and a few hundred seconds for ions (Baumjohann & Treumann, 1997, page 35). The trapped particles also experience two separate azimuthal motions called the 'curvature' and 'gradient' drifts, each of which entails both an eastward electron drift and a westward ion drift. The curvature drift arises since the curvature of the field lines with latitude causes a centrifugal force on the particles, perpendicular to both the magnetic field and its direction of curvature (Baumjohann & Treumann, 1997, page 22).

The gradient drift is caused by the decrease in magnetic field intensity with distance away from the Earth. The charged particles gyrating about the magnetic field lines experience a tighter gyration curvature on the side of their orbit with the stronger magnetic field – this causes a net motion, perpendicular to both the magnetic field and the direction of its change in intensity. The combined gradient- and curvature-driven azimuthal drift is of much longer period than the bounce motion, typically a few hundred hours. Most charged particles do not remain trapped long enough to fully orbit the Earth, though highly energetic particles can perform closed orbits (Baumjohann & Treumann, 1997, page 39). The relative motion of the ions and electrons (due to the curvature and gradient drifts) together causes a westwards ring current which encompasses the Earth (Daglis et al., 1999; Kivelson & Russell, 1995). The ring current is the major source of signal in a geomagnetic storm, the name given to a magnetospheric disturbance caused by either a prolonged southwards orientation of the IMF or an increase in the solar wind magnitude (Kivelson & Russell, 1995, page 406). Geomagnetic storms have three phases: the initial phase (sudden storm commencement) lasting at most few hours, a main phase that lasts a good fraction of one day, and a recovery phase which typically lasts several days. The initial phase is characterised by a sharp increase in the horizontal magnetic field component on the dayside of a few tens of nT, caused by the compression of the magnetic field lines by the increased solar wind intensity. During the main phase, the depression of the horizontal component of the magnetic field can reach up to 250–500 nT at the equator (Kivelson & Russell, 1995; Campbell, 2003, page 407, 139 resp.). However, storms are infrequent (occurring typically less once per month, dependent on the mean solar activity level), and in this thesis I focus mainly on times of quieter activity, during which the ring current has a magnitude of roughly 15 nT, modulated (as I discuss in chapter 5) by the 11-year solar cycle (Lühr & Maus, 2010).

The magnetosphere extends some 200 R_E (Campbell, 2003, page 139) in the anti-sunwards direction (24:00 LT), forming the magnetotail. At a distance of 10 R_E from the Earth (and further out), a cross-tail current (neutral sheet current in Figure 1.4) flows in the dawn-to-dusk direction, with a magnitude of about 8.5 nT southwards (Lühr & Maus, 2010). The cross-tail current completes its circuit via a return path across the tailside magnetopause (labelled

tail current in Figure 1.4). At the opposite LT (12:00) on the sunwards side of the Earth, the magnetopause (Chapman-Ferraro) current has a stand-off distance of typically 11 R_E. As it is also directed dawn-to-dusk, the magnetic effect at Earth is in the northwards direction with a magnitude of 25 nT on average (Campbell, 2003, page 137). The field-aligned currents (FACs) shown in Figure 1.4 connect magnetospheric current generators to the high-latitude ionosphere. An intensification in the dusk-side ring current (called the partial ring current) forms a circuit along these field lines. The FACs generate a toroidal magnetic field which affects the Earth primarily at high latitudes, and to a certain extent also at mid and low latitudes, though this is presently not well understood as not all components of the contributing currents are adequately resolved (Sabaka et al., 2004).

Compared to the magnetosphere, the ionosphere is a much thinner shell of conducting plasma and is host to much weaker electric current systems. However, the currents' increased proximity to ground-based observatories and LEO satellites means that the associated magnetic fields can have apparent magnitudes comparable to or greater than the quiet-time magnetospheric signals. Figure 1.6 is a schematic cross-section of the Earth and near-Earth space environment (with a quasi-logarithmic radial scale), and shows the major ionospheric current systems which are important in this thesis. The ionospheric plasma is created principally by solar UV (ultra-violet) and X-ray radiation dissociating the neutral atmosphere at altitudes between 90 and 600 km. The resulting ions and electrons are entrained in the ambient magnetic field lines, and are subject (more so the ions) to forcing from the primarily thermally-driven flow of neutral atoms in the ionosphere (called the neutral wind). The neutral and electrodynamic forcing of the spherical shell of ionospheric plasma – in an environment of curving magnetic field lines – combines with the diurnal pattern of atmospheric dissociation and recombination to create a complex array of divergences in the electric charge distribution. Some of the resulting magnetic fields are shown in Figure 1.6, and I describe them here, leaving a more detailed description of the divergences to chapter 3.

The major ionospheric current system is one of hemisphere-scale vortices at mid-latitudes

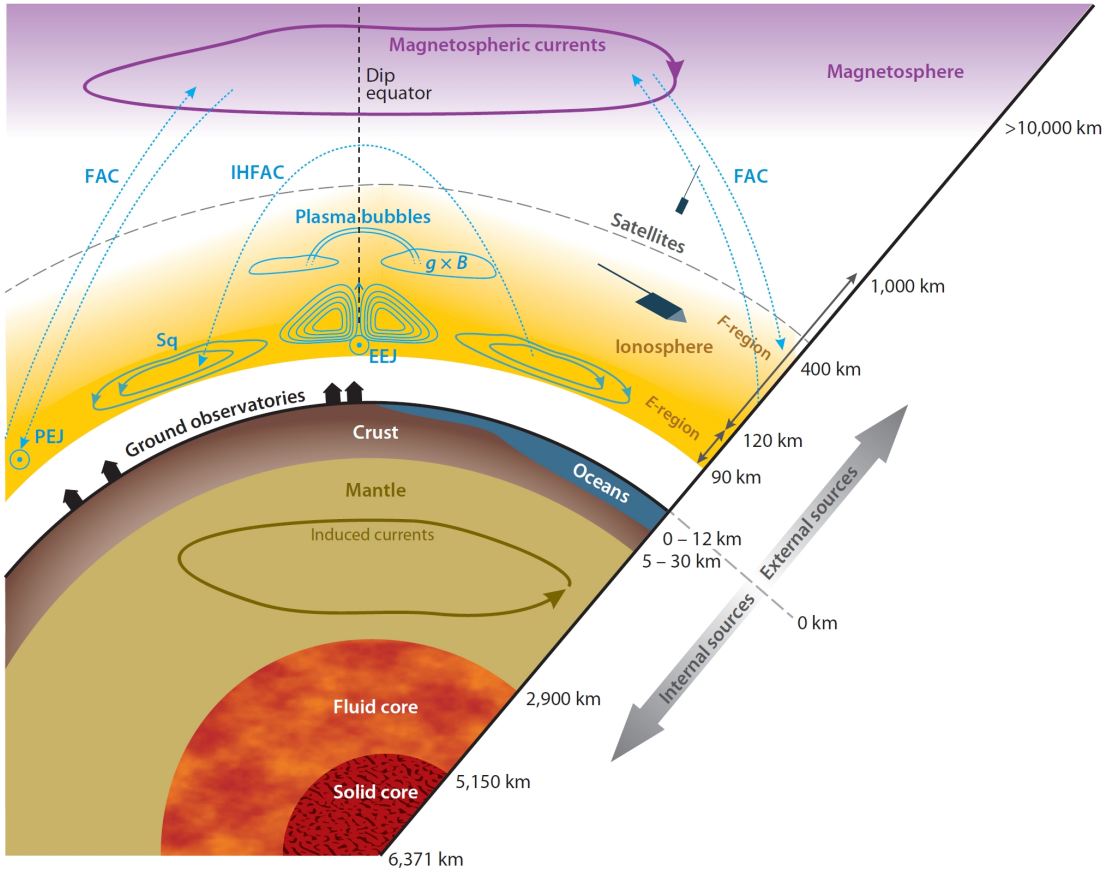

Figure 1.6: Schematic cross-section of near-Earth magnetic field contributions. Here B is the regional magnetic field vector; EEJ, equatorial electrojet; FAC, field-aligned currents (with IHFAC the interhemispheric field-aligned current); g, Earth's gravity vector; PEJ, polar electrojet (also called auroral electrojet); Sq, solar quiet. Image and description from Olsen & Stolle (2012).

on the sunlit side of the Earth. The diurnal variation of the magnetic field of this system is called the solar-quiet (Sq) signal, and the same name is given to the current vortices. Sq can produce a diurnal disturbance of around 70 nT on the Earth's surface at mid-latitudes, though values of ± 40 nT are more typical (Sabaka et al., 2004). The Sq currents peak at an altitude close to 110 km (Heelis, 2004; Kelley, 2009), hence there are substantial volumes of plasma in the altitudes above the current vortices. Any divergence in the currents produced in each hemisphere can thus be balanced by inter-hemispheric field-aligned currents (IHFACs) flowing in the upper ionospheric plasma. Both the large-scale magnetospheric FACs and the ionospheric IHFACs cause toroidal magnetic fields which impinge upon LEO satellite sampling shells.

At the dip equator there exists a strong eastwards-directed current called the equatorial electrojet (EEJ), which has a magnitude of about 100 nT in near-equatorial magnetic observatory records (Campbell, 2003, pages 7-8). At higher latitudes there exist strong and very spatiotemporally variant magnetic effects resulting from connection of the magnetosphere to the ionosphere via the FACs, creating the auroral electrojets. In the auroral regions, the magnetic fields resulting from this interaction of the ionosphere and magnetosphere can have magnitudes of 1000–2000 nT during geomagnetic storms (Backus et al., 1996; Campbell, 2003).

The Earth rotates underneath the Sun-synchronous magnetic field structures in the ionosphere and magnetosphere, producing a LT and UT dependence on the magnetic signal recorded at a geographic location. I will discuss the effect of the external fields on internal field estimation throughout this thesis – in the next section I summarise some of the temporal scales of field variation to put the overlap of the internal and external field variations in context.

1.2.3 Temporal variations in the geomagnetic field

Characterising the processes which create the internal field is the only way to forecast changes in the magnetic field on long time periods. As discussed by Backus et al. (1996), the temporal variations of the geomagnetic field span a wide range of timescales, and have high magnitudes even at the most rapid fluctuations measurable. However, only some of these variations are important for studies of the core field and the processes underlying its generation. At periods of 10^{-15} s there is a magnetic field related to sunlight, which on the sunlit side of the Earth's surface has a (mean) magnitude of 2400 nT (Backus et al., 1996, page 9). At slightly longer periods, disturbances from lightning strikes cause waves in the ionospheric plasma which propagate along field lines to a conjugate point in the opposite hemisphere – these waves have a period of 10^{-3} s. Both the oscillations discussed so far are much too rapid for measurements aimed at characterising internal field processes, since the data used for this (*e.g.* from ground-based geomagnetic observatories – discussed in more detail in section 1.3.1) are typically sampled at (*i.e.* averaged over) one or more seconds (Campbell, 2003, page 221).

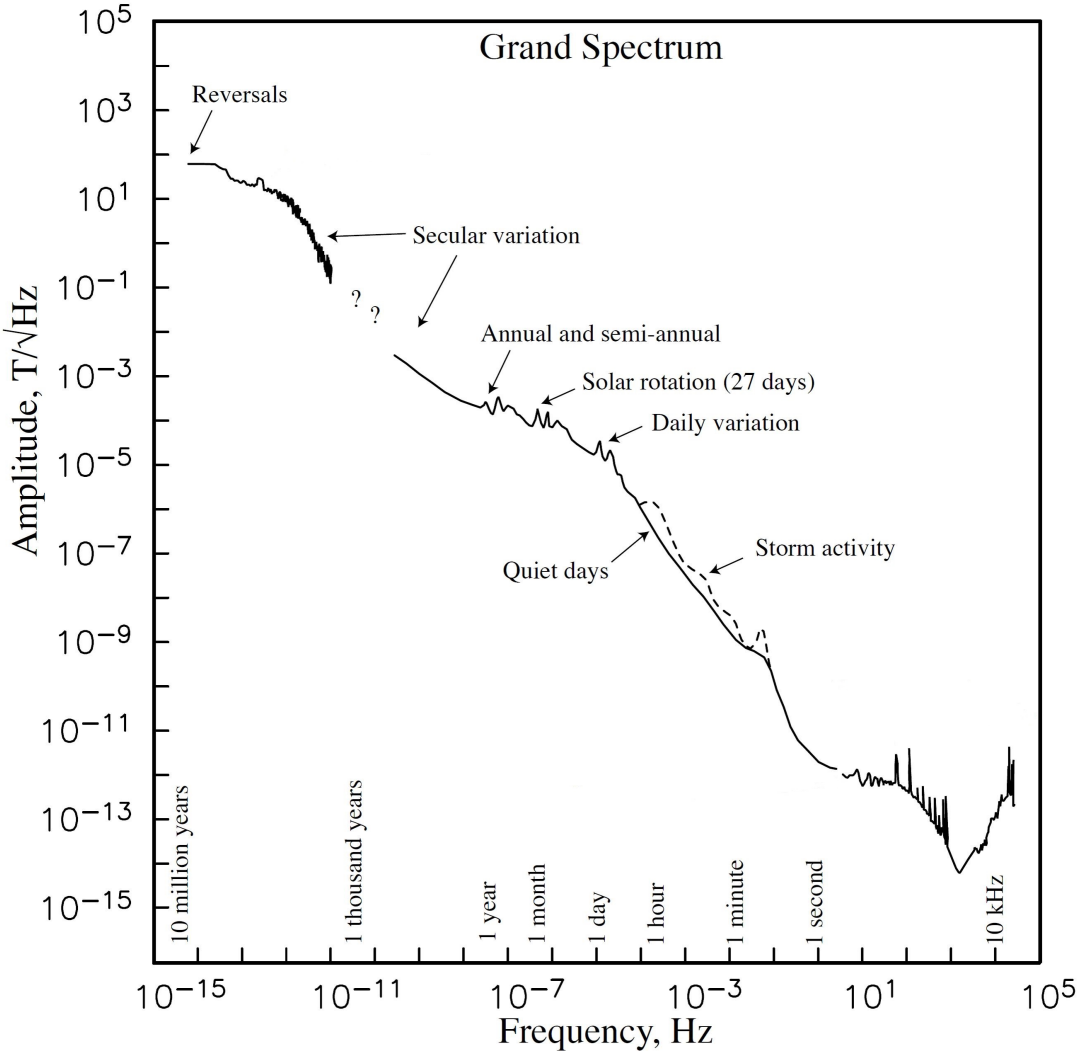

Figure 1.7: Temporal spectrum of geomagnetic field variations, with amplitude shown as a function of frequency. After, Constable & Constable (2004), showing the timescales of interest in this thesis.

The temporal variations of interest in this thesis are shown in Figure 1.7. External field disturbances dominate the temporal spectrum on timescales of seconds to months, and thus have a strong influence on attempts to model the internal field. On time periods of months to centuries, the SV of the internal field is dominant, though the modulation of the external fields by the 11-year solar cycle provides a further complication to this overlap of internal and external field signals. Solar cycle effects on the external fields are discussed in detail in chapter 4. As stated by Beggan (2009), the chaotic reversal of the geomagnetic field polarity on the scale of hundreds of thousands of years places somewhat of an upper bound on the timescales on which inferred core flow can be used to forecast the magnetic field variation. Thus the term SV has historically been used to refer to changes on the magnetic field on the timescales of years to centuries. In this thesis I rely on the wealth of data collected during the Decade of Geopotential Field Research (DGR) (*e.g.* Friis-Christensen et al., 2009). I am therefore interested mainly in SV that occurs on the decadal scale and shorter, and in the external field processes affecting internal field measurements within that time-range.

1.3 A history of magnetic field measurement and data

1.3.1 Development of the modern observatory

In this section I will describe the data sources I use, in addition to placing their existence in a historical context. It has been known for roughly 4000 years that certain materials have magnetic properties, as the inhabitants of early China recognised that naturally magnetised magnetite would attract iron (Kivelson & Russell, 1995; Backus et al., 1996). These materials were termed 'lodestones' (meaning leading stones) by later Western writers. The use of magnetic materials for navigational purposes has its earliest record at 250 BCE in China (Campbell, 2003). The increasing use of magnetic measurements in the worldwide exploration of European navigators allowed William Gilbert, in 1600, to publish *De Magnete*, the first textbook on global geomagnetism, and to infer that the Earth behaved as "a great magnet" (*e.g.* Figure 1.8). In 1635 Henry Gellibrand completed the study of Edmund Gunter on temporal variations in declination (the horizontal angle between geographic and magnetic north), pub-

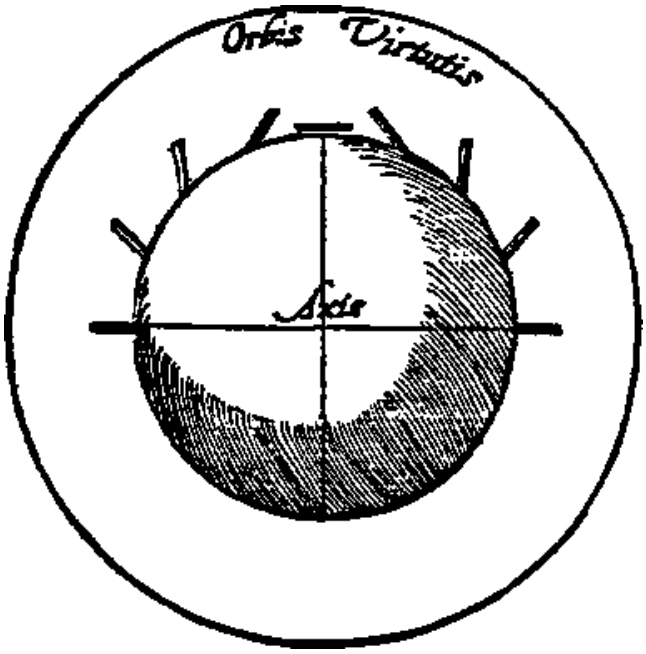

Figure 1.8: A early depiction of the Earth's dipole-dominated magnetic field, from Gilbert's De Magnete. Gilbert used a spherical lodestone to infer that the Earth itself behaved as a magnet. Image from Akasofu & Kamide (1987), but accessed online at `http://www-ssc.igpp.ucla.edu/ssc/tutorial/magnetosphere.html` on 19/05/2013.

lishing the discovery of SV (Backus et al., 1996). In 1680, Edmund Halley published a model for the westwards drift of declination features consisting of the westwards motion of dipoles deep within the Earth, analogous to the modern understanding of the SV (Backus et al., 1996).

It was in 1832 that Karl Friedrich Gauss combined existing methods for measuring relative magnetic field intensity with his deduction of the dipole field geometry to produce the first measurement of the absolute (horizontal) intensity of the geomagnetic field (Malin, 1982; Langel, 1987). Gauss later invented spherical harmonic analysis (discussed in more detail in section 1.4), showing that the field sources are predominantly internal to the Earth (Backus et al., 1996). Gauss was instrumental in beginning the development of a global network of magnetometer stations (an effort which von Humboldt later expanded on) providing the continued time series of magnetic recordings on which most modern efforts to study the magnetic field depend (Kivelson & Russell, 1995).

The number of worldwide geomagnetic observatories has been increased by international scientific campaigns such as the International Polar Years of 1882/3 and 1932/3, and the International Geophysical Year in 1957/8 (Gubbins & Herrero-Bervera, 2007). There are around

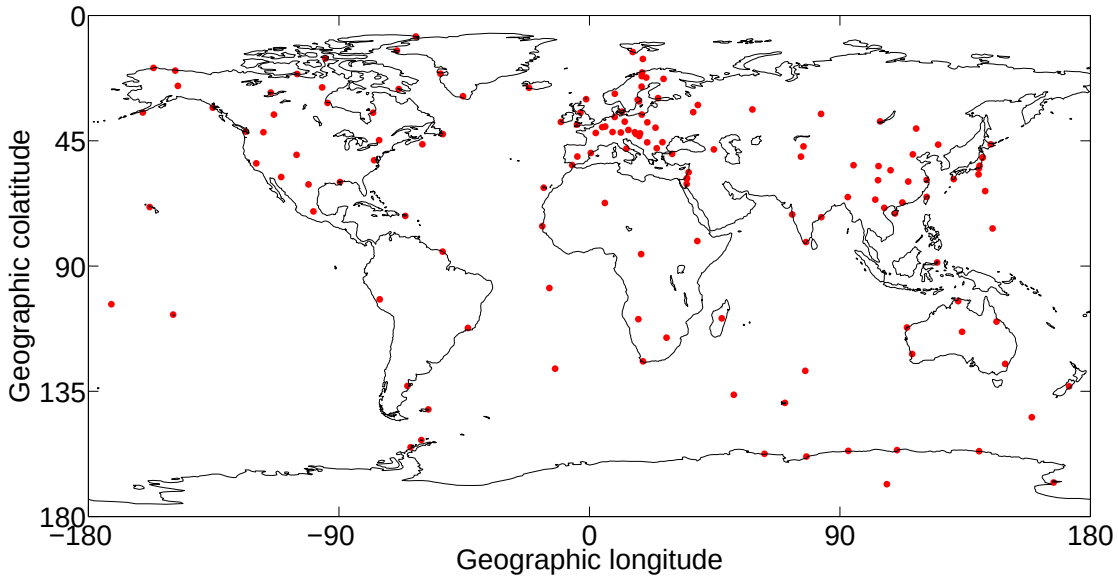

Figure 1.9: Geographic distribution of 149 magnetic observatories which each contributed several years of hourly mean vector magnetic data within the period 1997-2010 (coverage is not temporally continuous for all observatories). From the study of Olsen et al. (2011).

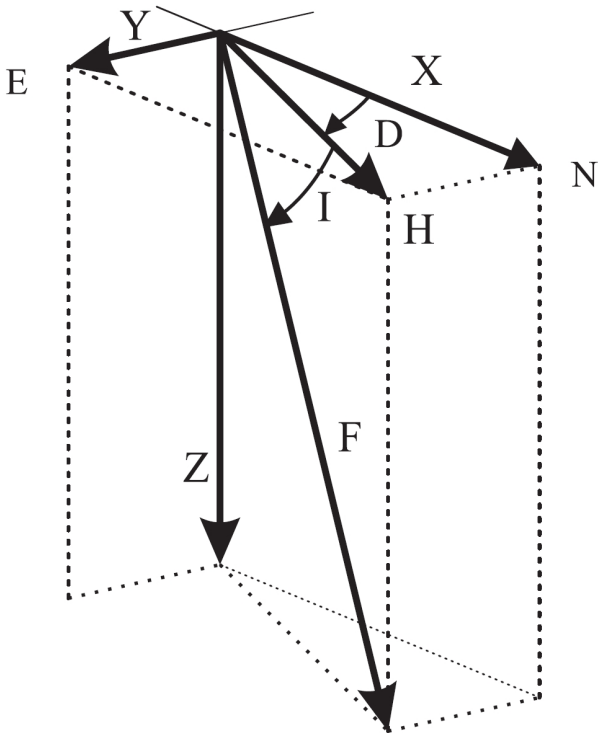

Figure 1.10: Standard geomagnetic coordinate terminology, depicting declination D, inclination I, horizontal component H, total field vector F and three orthogonal components X, Y and Z. N, E denote North, East respectively. From (Glassmeier et al., 2008).

180 currently operating observatories, many of which operate to internationally-defined standards such as those of INTERMAGNET (International Real-time Magnetic observatory Network). A magnetic observatory conforming to these standards will provide minute-means of magnetic vector data with a resolution of 0.1 nT (Campbell, 2003, page 221), though 1 Hz data are becoming more common. Historically, magnetic observatories have used a combination of absolute (scalar) and variometer (vector) instruments to express the magnetic field vector in an Earth-centred, Earth-fixed (ECEF) geographic coordinate system, with the three measured components being the total intensity (F in Figure 1.10), the declination (D), and either the horizontal intensity (H) or inclination (I) as the third value (Jankowski & Sucksdorff, 1996; Auster et al., 2007). The orientation of the magnetometer is provided by astronomical observations (Jankowski & Sucksdorff, 1996). Owing to the long history of magnetic field observation, a broad range of instrument types have been employed in obtaining these measurements. Modern observations are provided by fluxgate vector magnetometers (with frequent separate absolute measurements) and are expressed as the three components X, Y and Z, respectively geographic North, East and radially downwards, as shown in Figure 1.10. Magnetic observatories provide valuable long time series of data, essential in studying deep-Earth processes. However, as is apparent from Figure 1.9, the geographical distribution of observatories has a limited coverage in oceanic regions, and also in the southern hemisphere. The best geographic coverage is obtained from satellite measurements.

1.3.2 The satellite era

In the years following the launch of the COSMOS-49 mission in 1964, there were several isolated periods of satellite-based measurement of the magnetic field, including the 1979-1980 Magsat mission (Campbell, 2003, page 224) which I refer to in chapter 3. However, spaceborne magnetometry (and the Decade of Geopotential Field Research) started in earnest with the 1999 launch of the Ørsted satellite (Neubert et al., 2001; Friis-Christensen et al., 2009), followed shortly by the launch of the CHAMP (Reigber et al., 2002), and SAC-C (Colomb et al., 2004) satellites in 2000. I use data from the CHAMP and Ørsted satellites in this thesis.

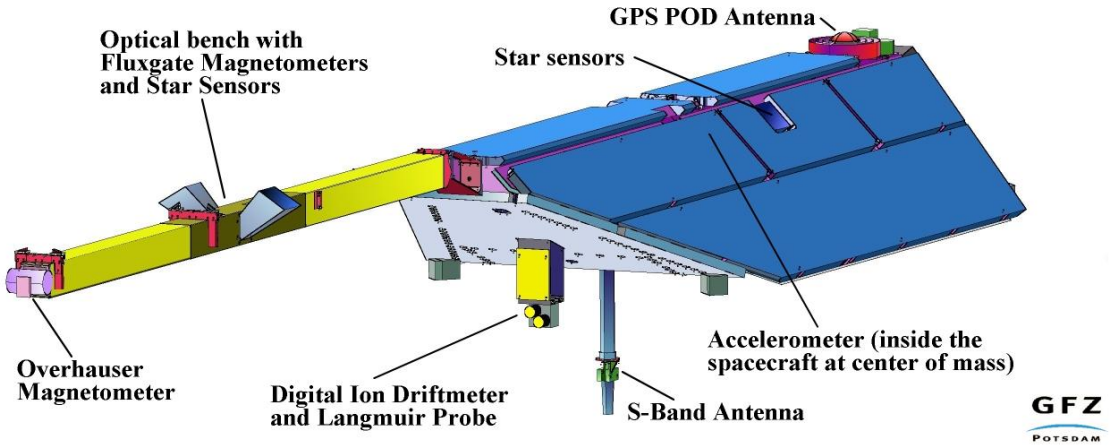

Figure 1.11: Schematic illustration of the design of the CHAMP satellite (Reigber et al., 2002). Note the design of the boom, with the scalar (Overhauser) and vector (fluxgate) magnetometers, and the two star cameras. These are discussed further in the text. Image from `http://op.gfz-potsdam.de/champ/systems/main_SYSTEMS.html`, accessed 09/05/2013.

Both CHAMP and Ørsted precess slowly in LT (described fully in chapter 3), and have each sampled the full 24hr LT span multiple times. The design of CHAMP and Ørsted – that of separate vector and scalar magnetometers acting together to provide an orientated vector magnetic measurement – provided the construction model on which the upcoming *Swarm* satellite mission (Friis-Christensen et al., 2006) was based. Uniquely, *Swarm* will sample multiple LTs simultaneously (this is discussed in more detail in chapter 2). Unlike ground-based observatories, satellites are continually moving with respect to the geographic coordinate frame, and the consistent orientation of the vector data into a ECEF frame provides somewhat of a challenge. Orientation of the satellite data is particularly important to the study of electric currents at satellite altitude (chapter 3), so I will summarise the orientation process below.

Orientation of satellite magnetic vector data is performed in-flight, at the time and location of each measurement. For an example, the design of the CHAMP satellite is shown in Figure 1.11, with the arrangement of the fluxgate (vector) and Overhauser (scalar) magnetometers shown in relation to the star sensors (two orthogonal cameras). The star sensor (also called a star tracker) is used to provide the orientation of the satellite, whilst the GPS antenna provides a position for the satellite. Satellites in LEO travel at speeds of 7-8 km/s, meaning that a small timing-error can lead to a large error in the positioning of the satellite measurement.

However, the use of GPS for position determination means that LEO positions are precise to cm, so along-track positioning errors are assumed negligible in this thesis. However, the data errors are anisotropic, and attitude error does affect the measurements, as I discuss in chapter 3.

The process of obtaining accurate vector components in the ECEF frame is referred to as the 'calibration' of the magnetometer readings, and as discussed by Olsen et al. (2003a, 2006b, 2007), involves the following key steps. At each measurement of the satellite, the scalar magnetometer reading is used to mitigate the 'drift' with time of the more sensitive vector instrument. The magnetic measurements are then said to be in the coordinate frame of the vector fluxgate magnetometer (the VFM frame). The three Euler rotation angles (Weisstein, 2003) required to translate the magnetic vector from the VFM to the star tracker (STR) frame are estimated prior to launch. However this alignment is liable to change at launch and during the satellite mission and so the Euler angles are additionally determined in-flight. The angles are estimated via comparison of the vector measurement with an independently-calculated internal field model, though Olsen et al. (2006b) describe a magnetic field model in which the Euler angles are coestimated in the inversion for the model coefficients. The coestimation procedure is a more stable approach than the comparison with an existing model and provides a less 'circular argument', but my studies are in fact dependent more heavily on the timescale on which the Euler angles are estimated. Whilst the satellite measurements are rotated from the VFM to the STR frame at each measurement taken, the Euler angles used in the orientation are estimated over many orbits of data. A long estimation timespan is desirable in order to mitigate LT-dependent effects on the magnetic vector alignment (since these satellite precess slowly in LT), but in the case of CHAMP this is not possible, as the instrument bench housing the VFM and STR is known to undergo thermally-driven bending (Olsen et al., 2006b). The same is likely true for Ørsted's bench, but a change in its orientation with time is not resolved. Thus whilst Ørsted's in-flight Euler angles are time-invariant, the angles for CHAMP's data must be estimated every few days. The calibration of CHAMP's data has the effect of aliasing signal which would otherwise be coherent in LT – I discuss the impact of the Euler angle estimation timescale for CHAMP in more detail in chapter 3. Following the calibration and rotation

stages, the orientation and position of the optical bench (given by the star tracker and GPS respectively) provide the information required to rotate the magnetic vector from the STR frame into the ECEF frame (Olsen et al., 2007). The calibration process for the magnetic measurements from Ørsted, CHAMP and *Swarm* involves the same essential stages as discussed above.

The orbital re-entry of the CHAMP satellite in September 2010[2], combined with the delay in the launch of Swarm to mid-2013 at the earliest, has left a gap in the otherwise continuous measurement of the Earth's magnetic field from space since 1999. Since the most accurate models of the internal field are presently achieved by combining the information from ground based observatories and satellites (*e.g.* Finlay et al., 2012), it would be desirable to obtain some data that could provide information in this epoch, and the C/NOFS (Communications/Navigation Outage Forecast System) satellite is a possible candidate. As described by Le et al. (2011), C/NOFS was launched in April 2008 in an orbit with 13° inclination from the geographic equatorial plane, a maximum initial altitude of 867 km and a minimum of 401 km. The satellite is equipped with a star tracker and a fluxgate magnetometer providing 1 Hz measurements, and orbits once every 97 minutes. I discuss the uses of alternative measurement platforms in more detail in chapter 5, including possible uses for the C/NOFS data – here I will discuss the calibration applied to the C/NOFS satellite.

Euler angles for the non-orthogonality and orientation of the VFM with respect the star tracker on the C/NOFS spacecraft are determined prior to launch, but C/NOFS does not have a scalar magnetometer. The in-flight calibration is approached by a series of comparisons with magnetic field models, as follows (pers. comm. R. Pfaff, 2013). Initially, gain factors (which mitigate temperature-related drifts in the magnetometer) and adjustments to the orientation of the vector axes are derived from comparison of (at present) three years of C/NOFS magnetic data with the POMME-6 (discussed in section 1.4) model (Maus et al., 2010) prediction of internal and external fields. The derived rotation angles and gain factors will be assumed constant for the entire mission, having been found to be relatively invariant with time. In addition to

[2] `http://op.gfz-potsdam.de/champ/`, accessed 09/05/2013

this mission-spanning calibration, a daily calibration procedure is followed, consisting of additive offsets applied separately to the three magnetic vector components (hence, also implying a rotation of the magnetic vector). The offsets are computed once per day, and are designed to minimise the difference between the magnetic vector and a model prediction. The model is the International Geomagnetic Reference Field (IGRF) (Finlay et al., 2010) with the Dst index (described below) added to the model's north magnetic component. The offsets are smoothed over time and linearly interpolated to each of the 1 Hz measurement locations before being added to the magnetometer records.

The C/NOFS calibration process implicitly assumes that the magnetospheric signal is zonally invariant in LT. Therefore, over time the calibration process will either remove or alias magnetospheric contributions which deviate from a torus-geometry in LT. In addition to oversimplifying the magnetospheric content of the measured signal, the C/NOFS calibration combines (*e.g.* thermal) instrument drift into the field measurements. The daily calibration procedure mitigates the impact of this on long timescales, but also removes long-period UT signal in the field measurements. The temporally-smoothed calibration procedure used makes external field studies using C/NOFS data difficult, and the use of additive offsets impacts the usefulness of the data (in its present state of calibration) for internal field studies. Whilst the CHAMP calibration process aliases small amounts of signal which is otherwise coherent in LT (due to the continual re-estimation of the Euler angles), the comparison with the scalar magnetometer makes the CHAMP/Ørsted calibration procedure substantially more accurate than the calibration applied to C/NOFS data. Using C/NOFS data is beyond the scope of my thesis, though I discuss in chapter 5 how it could be processed for future work.

1.3.3 Magnetic field indices

Magnetic field indices are time series of magnetic field recordings which have been averaged spatially (typically across all longitudes and specific latitudes) and temporally (typically over one or several hours) in order to represent the time variations of a particular current system.

Such indices were originally derived from a limited number of observatories to provide a global overview of the disturbance level of the field before modern processing techniques became commonplace (Campbell, 2003). All indices must selectively 'misrepresent' the field in some way or other, but the aim in doing so is to offer more global information than that provided by a time series of magnetic field readings at a single observatory. As various additional magnetic field phenomena were identified, more precise indices were developed in an attempt to isolate the magnetic effect of specific current systems, but older indices have avoided updates to their methods of derivation to ensure historical continuity in the time series they provide (Kivelson & Russell, 1995; Perrone & De Franceschi, 1998). Indices are still used in field modelling today (Maus et al., 2006b) for the purposes of both data selection and external field parameter dependence (dynamic scaling, such as that applied in the C/NOFS calibration), and can thus be thought of as 'data'. I discuss the requirement for new indices, mainly due to issues in the computation of baselines, in chapter 5, but the magnetic indices I use in this thesis are two of the oldest: Dst and Kp (defined below), which I apply mainly in order to specify epochs of low geomagnetic activity in chapter 4.

The Dst (disturbance storm-time) index was designed to represent the symmetric equatorial ring current during storm-times. The early understanding of the effects of a geomagnetic storm was that of a simultaneous global depression of the horizontal (H) component (shown in Figure 1.10), attributed to the rapid intensification and slow decay of the ring current. Dst consists of the global average H component field strength of four near-equatorial observatories (Campbell, 2003). This average is corrected for core field SV and external field diurnal variation (Sq). SV effects are reduced by subtraction of a polynomial fit to midnight H values at each observatory on the quietest days in each year (Kivelson & Russell, 1995), which will also remove long-period external field variations. From this residual, the low harmonics of a Fourier analysis are used to define and subtract the Sq variation. The Dst index is the instantaneous global average of these adjusted residuals, and is given hourly.

From recent advances in magnetic field research, Dst is known to misrepresent the symmet-

ric ring current. Xu (2008) showed that the contribution of the symmetric ring current to the Dst index is only around 27%, with the remainder composed of the partial ring-, magnetotail-, and near-Earth cross-tail-currents. More recent research by Lühr & Maus (2010) (in addition to my results in chapter 4) has indicated that the baseline of Dst changes significantly with the solar cycle. Furthermore, Campbell (2003) pointed out that the correction for Sq does not account for changes in the diurnal variation with field activity, and hence Dst is "an ensemble of magnetospheric and ionospheric fields detected at middle and low geomagnetic latitudes". However, these drawbacks are well-known to the community.

The Kp index is a 'range' index (in contrast to the Dst 'mean' index), derived from the amplitude of variance of the most active of the two magnetic components D and H, in consecutive 3-hourly UT (Universal Time) bins. There are 13 mid-low magnetic latitude contributing observatories (Perrone & De Franceschi, 1998). Kp amplitudes are calculated relative to the diurnal variation, eliminating long-period departures of the field and resolving field variations of period close to that of the size of the bin (Campbell, 2003). The Kp index is more attuned to ionospheric variations than Dst due to the higher magnetic latitudes of the contributing observatories, and is used in chapter 4 to provide information on a decomposition of the external magnetic fields.

1.4 Magnetic field modelling approaches

A common method of producing a model of the magnetic field is to apply a spherical harmonic analysis (SHA), as described in (*e.g.*) Campbell (2003). SHA (as stated in section 1.3.1) was initially developed by Gauss in 1838, and is a method of representing the variations of a potential field via an expansion of a series of basis functions defined on the surface of a reference sphere. SHA solves Laplace's equation in spherical polar coordinates, relating the magnetic field measurements $\mathbf{B}$ to the Laplacian potential V, such that $\mathbf{B} = -\nabla V$. Under the assumption that there are no significant currents flowing through the measurement region, this expansion is a good analogue for magnetic field measurements made on the surface of the Earth, and hence SHA is widely used in geomagnetism – I discuss and make use of it in many of the following

chapters.

The scalar potential V is computed by SHA as follows (*e.g.* Olsen & Mandea, 2007)

$$\begin{aligned} V(r,\theta,\phi) \quad &= a \sum_{n=1}^{N} \sum_{m=0}^{n} \left(g_n^m \cos m\phi + h_n^m \sin m\phi\right) \left(\tfrac{a}{r}\right)^{n+1} P_n^m \\ &+ a \sum_{n=1}^{N} \sum_{m=0}^{n} \left(q_n^m \cos m\phi + s_n^m \sin m\phi\right) \left(\tfrac{r}{a}\right)^{n} P_n^m \end{aligned} \tag{1.1}$$

where a = 6371.2 km is a reference radius, r is altitude above a, θ and ϕ are geographic co-latitude and longitude, $P_n^m(\cos\theta)$ are the associated Schmidt semi-normalized Legendre functions (Langel, 1987), n and m are the degree and order of the expansion up to a maximum degree of N, g and h are the internal field Gauss coefficients, q and s the external field coefficients. The associated Schmidt semi-normalized Legendre functions, multiplied by the sinusoidal functions of ϕ, define the spherical harmonics – the basis functions of the expansion. The Gauss coefficients are the amplitude coefficients for these harmonics. As I discuss in chapter 2, the Gauss coefficients are solved for via least-squares (Menke, 1989). The orthogonality of the spherical harmonics allows for the Gauss coefficients to be determined independently of each other – increasing N in equation (1.1) does not alter the coefficients determined from a more truncated expansion (*e.g.* Langel, 1987). In practice, the retention of the orthogonality property is only perfect if the distribution of data about the reference sphere is ideal, necessitating that the inversion for the coefficients be regularised (smoothed) to avoid dependence on N. I discuss regularisation in more detail in chapter 2. Gauss used the property of the SHA (evident in equation (1.1)) which separates magnetic field contributions from current sources above and below the reference sphere, to show from the relative magnitudes of the internal and external Gauss coefficients that the magnetic field of the Earth is primarily internal in origin. The rapidity of the convergence of each set of internal and external Gauss coefficients (in effect, how the measurements fit into the basis functions) depends on the coordinate system used in the expansion. The best choice of coordinate system may be different for internal and external field representations. SHA is the most widely-used method for modelling internal magnetic fields, although it has some shortcomings that I mention at the appropriate points in

my research chapters. I make frequent use of existing magnetic field models in my research, and will describe the two most sophisticated examples below.

Spherical harmonic field models can be grouped into two broad camps: those that coestimate all field sources in a simultaneous inversion for the full geomagnetic field, and those field models which process or filter the data prior to modelling, in order to use the residual signal to represent some single desired 'part' of the field, for example, the Sq field. These two different modelling approaches are termed 'comprehensive' and 'serial', respectively. Note that a serial model may be made up of several discrete coestimations of different parts of the field: the boundary between the two approaches is not well defined. The comprehensive approach to field modelling mitigates errors resulting from *a priori* assumptions about the geometry of the external field, since the spatial structure of the source regions is determined in the inversion (Olsen et al., 2010a). However, as discussed in chapter 2, the geometry of the external fields must be known well enough in order to sample them effectively, or biases will occur in the model representation.

The most advanced model in the comprehensive series is the CM4 model of Sabaka et al. (2004). In the inversion for the CM4 model, the positions of the core, lithospheric, induced, ionospheric and magnetospheric fields are explicitly considered with respect to the locations of the sampling shells of both ground-based observatories and satellites. In this way, the contributions from the various field source regions are separable. CM4 provides a consistent representation of the internal fields to spherical harmonic degree 65, and external fields to degree 2, with a timespan from 1960 to mid-2002.

The POMME model (Maus et al., 2005, 2006b, 2010) is the most complex serial model of the near-Earth magnetic fields, though it has no parameterisation of the ionospheric fields. The authors use the approach of Maus & Lühr (2005), in which otherwise-complex temporally-varying signals in the geographic frame can be expressed (dependent on field activity) as static constant fields in a Sun-synchronous coordinate frame. Whilst CM4 adopts a similar approach

to its external field representation, it uses a shorter timespan of satellite data to constrain the external fields in LT. The modular arrangement of POMME allows for a relatively easy update process as new satellite data are acquired. In this thesis I make use of the CM4 model rather than the POMME model since the CM approach results in a more representative partitioning of the bulk signal amongst the physical sources, in a weighted least squares sense.

1.5 Aims of the project

In this thesis I study the external fields in order to better understand their effects on efforts to characterise the magnetic field internal to the Earth. The variations of the external fields have long been considered a source of noise when defining the internal field. Thanks largely to the abundance of high-quality satellite magnetic data collected during the DGR, it has recently become clear that the spatial and temporal variations of the external fields are the primary sources of error in efforts to model the spatial and temporal variations of the internal field. It is typical, when constructing a model of the internal field, to select only the very quietest magnetic field conditions in order to limit the impact of the contributions from the external fields. A strong implicit assumption in this approach is that when the external fields (in the reduced data set) are averaged over a few days or more, the external field variations in the time series will tend towards a zero mean (albeit with a much-reduced magnetospheric contribution still present) and have a Gaussian or Laplacian error distribution. If the external fields have variance or features which are not reduced or removed by this selection process, then the internal field model could be biased. The impact of these biases and how commonly they occur in internal field modelling efforts are discussed in a later study (chapter 2), and form one of the main threads of discussion in this thesis, so I will not summarise them here. However, the following brief discussion of the causes of the misrepresentation of the external fields will help place the aims of this thesis and the rest of this background chapter in better context.

The major reason that the external fields are the largest source of error in internal field modelling is that they are organised most strongly in a Sun-synchronous coordinate frame, whereas the internal field is more Earth-synchronous. When choosing a set of measurements

to represent the internal field, the spatial distribution of the data is understandably optimised for resolution of fields which are Earth-synchronous, meaning that the representation of the external fields is simplified. This trade-off affects some sources of data more than others. The satellite data collected during the DGR have provided a geographic coverage far superior to that of the permanent observatory network, but at the expense of an adequate simultaneous coverage in LT. Since the external fields vary strongly in space and time, when using satellite data to model the internal field it is very easy to alias even small-magnitude external field contributions into the internal field model (a process discussed in detail in chapter 2). The application of quiet-time data selection criteria reduces the contribution of the external fields considerably, but not totally. The following (non-exhaustive) list offers some scenarios for the introduction of external field biases during quiet times.

- The extent to which the external fields average to a zero or consistent mean is heavily dependent on the manner in which they are sampled.
- There exist external fields which are systematically present, yet which are unmodelled and hence not removed in the quiet-time selection procedure.
- The indices used to specify quiet-times have baselines which are known to have a varying error at different points in the solar cycle, and which are known to be non-applicable in regions outside the mid-latitudes.

All these issues can be improved upon by knowing more about the external fields, so that efforts to select (and eventually correct) for their disturbances are more successful. At present a substantial proportion of the satellite data collected during the DGR have limited use in internal field modelling as the external field contributions are too strong, or not understood well enough to be able to mitigate their effects. Despite this, the satellite data collected during this period have provided some of the most accurate models of internal field variations yet (Finlay et al., 2012). The reliance on satellite data for magnetic models is set to increase, and in order to improve the description of the external fields in tandem with this development, a more effective sampling basis for the external fields is required. The European Space Agency's *Swarm* constellation of three satellites (described in Friis-Christensen et al. (2006), with further details

provided in chapter 2) is due for launch in late 2013, and is expected to provide magnetic data from LEO with unprecedented accuracy and precision. As I discuss in later chapters, satellite constellation missions are crucial in providing the required simultaneous sampling of a representative number of LT sectors, whilst also providing near-total geographic coverage.

In the following chapters I present a series of studies which aim to improve the description of the external fields using constellation satellite data, in order to better identify systematic biases and errors resulting from incorrect treatment of external field measurements. There is an emphasis on methods for separating external from internal magnetic fields, such that the relative magnitudes of each external field source can be understood in terms of its impact on internal field estimation, at different points in the solar cycle. The distribution of constellation data is not ideal, so a key part of these studies is to identify the shortcomings of the constellation data and to identify how other sources of information can be used to supplement them - this is important given the increasing contribution of satellite data to internal field studies. My aim of improving the description of the external fields implies not just measurement and characterisation, but also an aim to improve upon or complement existing magnetic field modelling techniques. I will adapt established mathematical methods for application to the problem of near-Earth external field resolution, in order to produce new data sets and new information.

My major research questions are as follows.

- To what extent are the external fields encountered in satellite magnetic data being oversimplified, and how can this be avoided in order to resolve the internal fields to the highest degree of accuracy?
- What is the impact of oversimplifying our description of the external fields on our ability to properly model the internal field and its change in time, particularly when using spherical harmonic analysis?
- How can satellite constellation missions help improve our description of the external magnetic fields, whilst benefiting from the advances they make possible?

I do not attempt to answer all the questions raised in the proposed field of research, but each individual study I present is intended to contribute to the increasing knowledge of the processes governing the Earth's external fields. I have focused on producing advances in the descriptions of the morphology of the external fields in LT, the impact of the fields on standard modelling practice, and their distribution at LEO altitudes. My focus is on systematic biases affecting internal field models. Since internal field models use magnetic data primarily from quiet times and mid- to low-latitudes, I research these conditions and regions in the most detail. Whilst I do not study the internal fields explicitly in chapters 3 and 4, I use the magnitudes of temporal variation of the internal field (discussed in more detail in chapter 2) to place my study of the external fields in appropriate context. Each of the chapters 2 to 4 covers one or more of the main research questions. Each chapter is titled after the principal mathematical method applied in its study, since the method in each study was chosen as the most suitable to answer the research questions posed.

In chapter 2, I assess the impact of aliased external field contributions in measurements used to model the internal field via spherical harmonic analysis (SHA). In particular, I assess the effect to which the constellation geometry of the upcoming *Swarm* mission is able to facilitate the separation of the internal and external fields in magnetic data, and to mitigate the aliasing effects. In this study I make use of the Virtual Observatory (VO) method of Mandea & Olsen (2006), which is used to process satellite data in order to approximate the monthly mean values of the magnetic field available from ground-based observatories. The VOs offer a superior geographic coverage to the ground-based observatories but suffer from temporal biases, as identified by Beggan (2009). The study of chapter 2 uses synthetic *Swarm* data and hence does not aim to resolve new external field information, rather to assess the benefits of utilising a constellation of satellites in internal field modelling. From this I was able to infer a set of general rules for adequate resolution of the external fields including a quantification of the error incurred when these conditions are not met, and the results of this study had a formative influence on the objectives for the remainder of my thesis.

In chapter 3, I use a pair of satellites (Ørsted and CHAMP) orbiting at different altitudes to estimate the average current density (via Ampère's integral) flowing within loops formed from discrete overflights of the two spacecraft. The initial motivation for this study was to assess (at the time) unexplained biases in the study of chapter 2, which I inferred might be due to in-situ electric current flow at LEO altitudes. From the results I infer the systematic effects these currents may be causing in internal field models. I compare the satellite current density estimates with predictions from the CTIP (Coupled Thermosphere-Ionosphere-Plasmasphere) Model, a 3-dimensional numerical model of ionospheric composition and temperature.

Finally, in chapter 4 I investigate the morphology of the (long-period) external fields in LT, in order to resolve the external field signals important to mantle induction. Mantle induction studies are a key part of the *Swarm* mission science objectives, and the *Swarm* constellation geometry is known to undersample the external fields, *i.e.* it does not sample enough LT sectors simultaneously. In this study I apply the method of Empirical Orthogonal Functions (EOFs) to ground-based observatory data. EOFs are composed of a series of vectors defining patterns which provide a decomposition of the spatio-temporal structures contained in the data, with partitions arising from the data themselves. The aim of this study is to overcome overly-simplistic assumptions about the inducing field morphology and to assess how important the loss of a complete coverage of LT sectors will be to *Swarm*. The successful application of the EOF method to long time series of data opens the possibility of a method to calculate new mantle induction transfer functions, used to probe the mantle conductivity distribution.

Chapter 2

An investigation of the properties of Virtual Observatories

2.1 Glossary

Table 2.1: Glossary of terms used in this chapter

Abbreviation	Explanation
VO	Virtual Observatory (plural VOs: Virtual Observatories)
SV	Secular Variation (temporal change of the core field, typically on scales of decades and centuries but here on the scale of month and years)
LT	Local time (not latitude-specific; does not account for seasonal variation)
B	Satellite magnetic vector data measurements (also called $\mathbf{B}_{sat}$)
SHA	Spherical Harmonic Analysis (likewise SH: Spherical Harmonic)
$\mathbf{B}_{MF}$	Core (main) field data predictions from a suitable SH model
$\delta\mathbf{B}$	Magnetic field residuals, after removal of a main field model
t	Indicator of time, used to denote a specific monthly mean magnetic field value
V	Laplacian potential
ESA	European Space Agency
Sq	Solar-Quiet (diurnal oscillation from ionospheric fields, in geographically-fixed measuring stations)
GEO	Geographic coordinate system (Hapgood, 1992)
LT frame	Description of spatial magnetic field variations, supplanting longitude with local time.
UT	Universal Time
SV frame	Description of magnetic field variations after the first time derivative is taken.
UT bias	Tendency of the VO sampling method to alias short period external field variations.
LT bias	Tendency of the VO sampling method to alias (effectively invariant) external fields in the LT frame.

Acknowledgements

The work in the Virtual Observatories chapter was done by me, with advice on associated theory and practice provided by my supervisors, who were all very patient in what turned out to be a rather slow-starting study. In particular I would like to thank Ciaran Beggan for Matlab assistance, and Kathy Whaler for spherical harmonic analysis theory assistance. I used Matlab code from Paul Leopardi, Ciaran Beggan, Kathy Whaler and Nils Olsen. My supervisors also helped with proofreading.

2.2 Motivation, objectives and background

2.2.1 Magnetic field sources and sampling coverage

The Earth's magnetic field is temporally and spatially variant on a wide range of scales. A detailed understanding of the complexities and dynamics of this system is required in order to make useful statements and predictions about the physical phenomena which give rise to its fluctuations. In turn, a better understanding of the interaction between the sources of the magnetic field allows a more effective delineation between desired and unwanted contributions in our measurements, and more accurate forecasting of magnetic field variations as a result. Any forecast of the magnetic field on time scales of months or longer must account for the change in the magnetic field resulting from the flow of liquid iron at the surface of the outer core. The first time derivative of the magnetic field contributions from the outer core is defined by this flow of iron, and is known as the secular variation (SV) of the magnetic field. Figure 2.1 shows the SV of the total magnetic field intensity, as described by the CHAOS-2 model of Olsen et al. (2009). The SV has a range of 180 nT/yr, equivalent to about 0.3 % of the range of the total field intensity of 60,000 nT at the magnetic poles (Campbell, 2003).

As discussed in chapter 1, the Earth's internal magnetic field interacts with the solar wind (plasma emitted by the Sun), creating the so-called external magnetic fields of the ionosphere and magnetosphere, with field-aligned currents coupling the two regions. Schematic images of the magnetosphere and ionosphere, illustrating their principal current systems, are shown

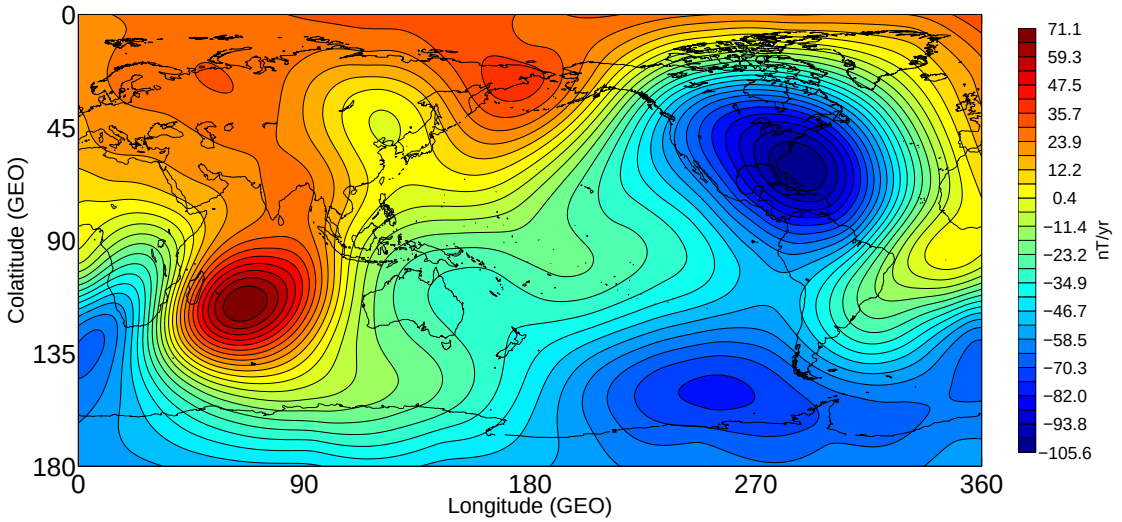

Figure 2.1: SV of total field intensity from the CHAOS-2 magnetic field model of Olsen et al. (2009), computed at the mid-point of August 1998 and August 1999, with spherical harmonic degree 14. Evaluated at 400 km altitude.

respectively in Figures 1.4 and 1.6 in chapter 1. In contrast to the SV, these external fields are strongly variant on time scales of as little as minutes, particularly in the auroral regions encircling the magnetic poles, where they are also much stronger (Campbell, 2003, page 143). In this chapter I will assess the bias in estimates of the internal field SV resulting from the external fields.

This study uses the geographic (GEO) system of spherical polar coordinates (r, θ, ϕ), where r is the length of the radial position vector from Earth's centre (and 'altitude' is the part of r above the reference sphere of radius 6371.2 km), θ (colatitude) is the angle between the radial vector and the Earth's rotation axis, and ϕ (longitude) is the angle in the equatorial plane between the radial vector and the Greenwich meridian (*e.g.*, Hapgood, 1992).

The magnetospheric fields are typically dominant in magnitude over the ionospheric fields, and since the magnetospheric fields are best assessed at mid-low-latitudes in the θ-component (south-pointing), this component is analysed in the greatest detail in this chapter. The θ-component of the SV is shown in Figure 2.2 and spans ± 70 nT/yr. I will later study situations in which the external fields are sampled such that they become highly aliased. In this

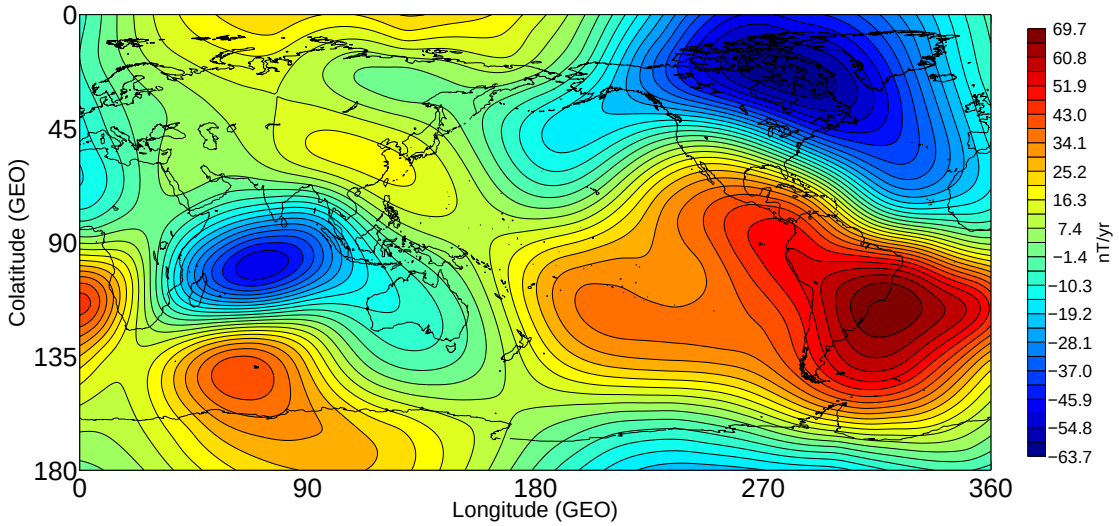

Figure 2.2: The θ-component of SV from CHAOS-2, for the same timespan as in Figure 2.1. Evaluated at 400 km altitude.

scenario, the magnitude of the external fields after the first annual difference is taken will be roughly equivalent to their static range. The ionospheric Solar-Quiet (Sq) magnetic fields have a θ-component magnitude range of about 70 nT (from the CM4 model of Sabaka et al. (2004)). At the Earth's surface at mid-latitudes, the magnetic signal of the magnetosphere is roughly double that of Sq, in periods of medium and low geomagnetic activity (Sabaka et al., 2004). The magnitude of the external fields can strongly bias internal field SV estimates if they are improperly accounted for.

The magnetospheric fields are highly spatially variant in their source regions (Campbell, 2003, page 148), but it is understood that at the Earth's surface they rarely exceed spatial wavelengths of degree-3 (around 13,000 km wavelength)[1]. The ionospheric fields are also approximately this smooth during quiet times, but can exhibit very small-scale structure (on the order of a few hundred kilometres) in the auroral regions during geomagnetic storms. Excepting these periods of high activity, the external fields are orientated with respect to the Sun. Therefore they are arranged in the local time (LT) frame – the reference frames used in this chapter are discussed fully in section 2.2.8, and will be important in later discussion. This chapter deals

[1] According to the conversion $(2R_E\pi)/n = \lambda$ where R_E is the Earth's radius, n is spherical harmonic degree, and λ is the wavelength corresponding to that degree (Glassmeier et al., 2008)

principally with magnetic fields at the altitude of satellites in low-Earth orbit (LEO), around 400 km. The geomagnetic field in this region also contains contributions from the static fields of the Earth's lithosphere, and induced fields from interaction between the Earth's conducting mantle, and the ionospheric and magnetospheric fields. As discussed in section 2.2.6, lithospheric fields are not important in this chapter, but the (time-varying) induced fields are treated as accessory to the external fields.

The SV is the largest magnitude change of the magnetic field which is sustained for periods of more than a year, and is important in studies of core flow processes. The spatial and temporal trends in SV are used as a basis for studying geomagnetic jerks (*e.g.* Olsen & Mandea, 2008) and the core field change on timescales of decades to centuries (Jackson et al., 2000). The data for estimating the SV are typically the monthly means of vector components from geomagnetic observatories. The observatories are geographically fixed, and rotate with the Earth underneath the LT-frame fields, recording a diurnal-period signal from the external fields. Imposed on this diurnal variation are the (typically stronger) random variations from the magnetospheric fields. When this combined signal is averaged over a few days or more, the external field variations in the observatory time series tend towards a zero mean, though a much-reduced ring current contribution is still present. When using these time series of data to estimate the SV, the monthly means are chosen because they reduce the external field contributions in this manner. The first time derivative of the monthly means is estimated by taking differences with an interval of 12 months, thereby removing the static fields (*e.g.* from the crust and lithosphere), as well as the annual (seasonal) variation. This process provides robust estimates of the SV, but can only be performed where permanent observatories exist. Figure 1.9 in chapter 1 is a map of the global distribution of permanent magnetic observatories. It can be seen that the relative paucity of stations in the oceans and in the southern hemisphere will lead to SV in these regions being more poorly constrained than (for instance) over Europe and North America. The impact of this irregular distribution differs dependent on what purpose the SV is used for. If the SV is used to study the morphology of the field's changes on long timescales in a global sense, spherical harmonic analysis (SHA) is commonly employed (SHA is discussed in detail

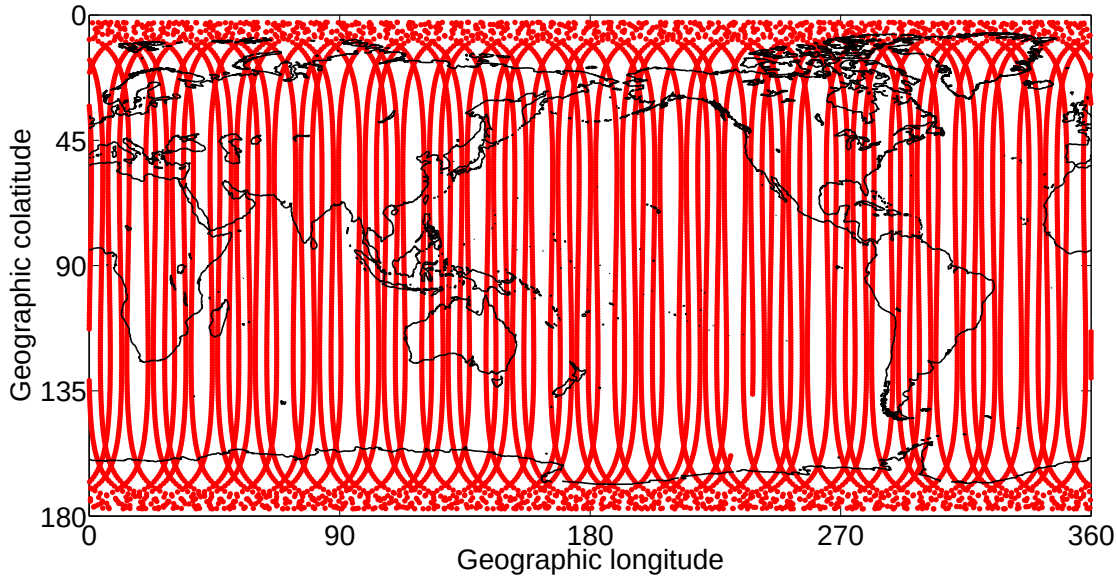

Figure 2.3: Geographic distribution of satellite data measurement locations for three days from a single satellite in a polar LEO with an inclination of 86.8°.

in chapter 1 and in section 2.3.1 here). However, the use of a SHA produces a globally-defined model, meaning that poorly constrained SV may contribute to the signal in regions of better-constrained SV. In studies of geomagnetic jerks, the time series of SV from a single observatory is used to determine the form of the jerk at that location. Jerks are spatially-localised phenomena which occur in different places at different times, and this localisation may span multiple observatories. However, in order to understand jerks more fully, a complete global coverage is ideal. Whether studying the SV at a specific geographic location, or globally, a more complete geographic coverage of data than that shown in Figure 1.9 is desirable.

Figure 2.3 shows the measurement locations from a single satellite, for three days of recording. Compared to the coverage in Figure 1.9, the satellite data have the potential to be an excellent resource for SV determination, if treated properly. Mandea & Olsen (2006) have described a method of using satellite magnetic vector data to resolve the SV at all geographic locations with a temporal repeat resolution of one month, in effect, calculating a series of monthly-mean equivalents from satellite data. The method produces what are termed 'Virtual Observatories' (VOs), and the benefits and drawbacks of this method will be discussed in detail in following sections.

2.2.2 The Virtual Observatory method

The VO method is designed to provide estimates of the monthly mean magnetic field at a fixed altitude and at a specific location in space, which represent the same components of the magnetic field as the monthly means from permanent observatories, but which are derived from satellite magnetic data. The localised field variations provided by the VOs can be used to provide estimates of SV at a geographically-fixed location almost anywhere on the globe, to mitigate the issues of incomplete global observatory coverage noted above. As discussed by Mandea & Olsen (2006), the VO method avoids certain difficulties in treating satellite data in a geographically-fixed analysis, namely that the satellite varies in altitude and geographic location, and that the fields being measured also vary temporally. The VO calculation process is described in Mandea & Olsen (2006), and can be summarised as follows. The satellite magnetic vector data $\mathbf{B}_{sat}$ have the predictions from a main field model (such as the CHAOS-2 model of Olsen et al. (2006b), which is summarised in section 2.3.1) removed from them at the location of each satellite measurement

$$\delta\mathbf{B} = \mathbf{B}_{sat} - \mathbf{B}_{MF}, \tag{2.1}$$

where $\delta\mathbf{B}$ are the magnetic vector residuals at each satellite measurement location, and $\mathbf{B}_{MF}$ are the main field model data predictions at a fixed date, thereby retaining the SV signal in the residuals. This step removes any changes in the static core field associated with the satellite altitude change. No data selection is applied to reduce external field contributions, in order to retain a similarity to the process by which ground-based observatory monthly means are derived. The VO monthly mean estimate is calculated from the $\delta\mathbf{B}$ which lie inside a cylinder (a data-bin) of radius 400 km about a specified centre-point. The radius of the cylinder is its horizontal extent rather than its height, and is a trade-off between spatial smoothing of the SV and the quantity of data binned – Mandea & Olsen (2006) tested cylinders of radius between 100 and 600 km and concluded that 400 km was most appropriate for resolution of SV. The data which lie inside each target region (at any altitude) throughout a period of a month

are converted to Cartesian coordinates (Hapgood, 1992), retaining the relative distribution of locations inside the data 'cloud'. The data positions (and the magnetic components) are then rotated to the 'target coordinate' position – this is the location of the resulting VO solution, computed such that an altitude of 400 km, and the geographic latitude and longitude of the data bin's centre-point, are at (0,0,0) in the local Cartesian target coordinate system. Mandea & Olsen (2006) discussed comparisons of SV obtained from 22 VOs with SV obtained from the same number of permanent observatories, in which the VOs were situated directly over their ground-based counterparts. The comparison of the two sets of SV (ground-based, and from VOs) showed reasonably good overall correlation between the two time series, but that there were substantial biases impacting the determination of the SV from the VOs, particularly in the horizontal components. In practice, the centre of the VO cylinder can be anywhere within the geographic span of the satellite's orbit, and as shown by Beggan (2009) the bin need not be a cylinder, nor of fixed radius.

It is assumed that the magnetic field residuals in the target coordinate system ($\delta\mathbf{B} = (\delta X, \delta Y, \delta Z)$) can be represented by a Laplacian potential V, such that $\delta\mathbf{B} = -\nabla V$. The VO solution process fits a smooth model to the data inside the VO bin, which neglects higher order derivatives by assuming a linearly varying potential field across the solution region. With these assumptions, V can be calculated using the parameters

$$\begin{aligned} V &= v_x x + v_y y + v_z z \\ &\quad + v_{xx} x^2 + v_{yy} y^2 - (v_{xx} + v_{yy}) z^2 \\ &\quad + v_{xy} xy + v_{xz} xz + v_{yz} yz, \end{aligned} \tag{2.2}$$

where x, y and z are the locations of the data in the target coordinate system and the subscripts denote partial derivatives. Calculating the VO solution from data collected over the course of a month reduces error from the temporal variations of the field, but still entails a non-Gaussian data distribution, with increased quantities of outliers from the unmodelled external

fields (the errors from which are not random). In the process of solving for the eight parameters making up V, the outliers must be downweighted, and so the parameters are estimated using iteratively reweighted least squares (IRLS) with Huber weights, minimising a one-norm measure of misfit. IRLS is discussed in Olsen et al. (2006b). A one-norm solution (*e.g.* Walker & Jackson, 2000) minimises the sum of the absolute values of the weighted residuals (of the model prediction and the original data). Huber weights (*e.g.* Walker & Jackson, 2000; Olsen et al., 2007) relate to the assumed probability density function (pdf) of the input data. A Huber pdf is a mixture of the Gaussian (normal) and Laplacian statistical distributions. The Laplacian pdf is similar in essence to the Gaussian, but with a slower fall-off of values away from the mean in the 'tails' of the distribution, and an elevated central peak. The Laplacian pdf therefore reflects the presence of increased outliers and anomalies in the data distribution. The Huber pdf is Gaussian in the centre and Laplacian in the tail. After the IRLS process, the estimate of the monthly mean magnetic field at the centre of the VO bin (the 'VO solution') is given by $\delta\mathbf{B} = -(v_x, v_y, v_z)$.

Typically, the VO solution is rotated from the target coordinates back into spherical polar coordinates and the $\mathbf{B}_{MF}$ value at the VO centre-point is added to the target-system solution. However, the main field addition is not required if the VOs are to be used for calculating SV, which is via

$$\mathrm{SV}_{VOmonthly} = \mathrm{VO}_{1month}(t+6) - \mathrm{VO}_{1month}(t-6), \tag{2.3}$$

where VO_{1month} is the VO solution for month t, and the units of t are months. The VO solutions contain the signal from static magnetic fields, including the lithospheric fields and any errors in the main field removal which are not related to SV. The calculation of the SV removes these static fields. As stated by Beggan (2009), a strong assumption in the VO method is that the external fields average to a zero mean over the course of a month, as is the case in the permanent observatory monthly means. This assumption does not always hold for the VOs and much of this chapter is devoted to discussing why.

The VO solution process automatically 'continues' the data to the reference altitude of 400 km, and limits error resulting from the spatial distribution of the data cloud, thereby accounting for both the altitude and geographic distribution of the satellite data. this is similar to the benefits offered by a spherical harmonic analysis (SHA, defined in chapter 1), though there are some key differences. From Campbell (2003), since the Laplacian function ($\nabla^2 V = 0$) follows from Maxwell's equations (Fleisch, 2008), and since these represent the physical behaviour of electromagnetic fields, a SHA determined on a valid reference surface defines the magnetic field everywhere. The VO local solutions account for local changes in data altitude, but do not provide a globally-differentiable function. Computing the SHA from VOs removes the benefits of the local solutions as far as jerk-studies are concerned, but allows the solutions to be extrapolated to the CMB, and to be inverted for advective core flow, as demonstrated by Beggan et al. (2009). In this chapter I make use of the internal-type SHA, and its use in conjunction with the VO solutions, in providing globally-defined magnetic field representations. The uses of SHA in this chapter are discussed in more detail in section 2.3.1.

2.2.3 Contamination in the VO solutions

An earlier study using VOs that I refer to frequently throughout this chapter is that of Beggan (2009). Rather than computing VOs overlying existing permanent observatories as Mandea & Olsen (2006), Beggan used two variants of the VO method to produce global maps of localised SV. In the first variant, 400 km-radius VO cylinders were spaced evenly at 10° intervals in colatitude and longitude, from 5° to 175° colatitude and -180° to 170° longitude, as shown in Figure 2.4. Noticing that this generated both gaps between bins, and overlaps of them, Beggan applied a second variant involving a tesseral partitioning of the sphere based on algorithms developed by Leopardi (2007), shown in Figure 2.5. In both the cylindrical and tesseral variants, there are 648 VO bins, and each bin acts as a half-space (Weisstein, 2003) with no upper altitude cut-off. I discuss the differences between these two variants later (section 2.3.1), but they are effectively equivalent and both are equal-area for all bins.

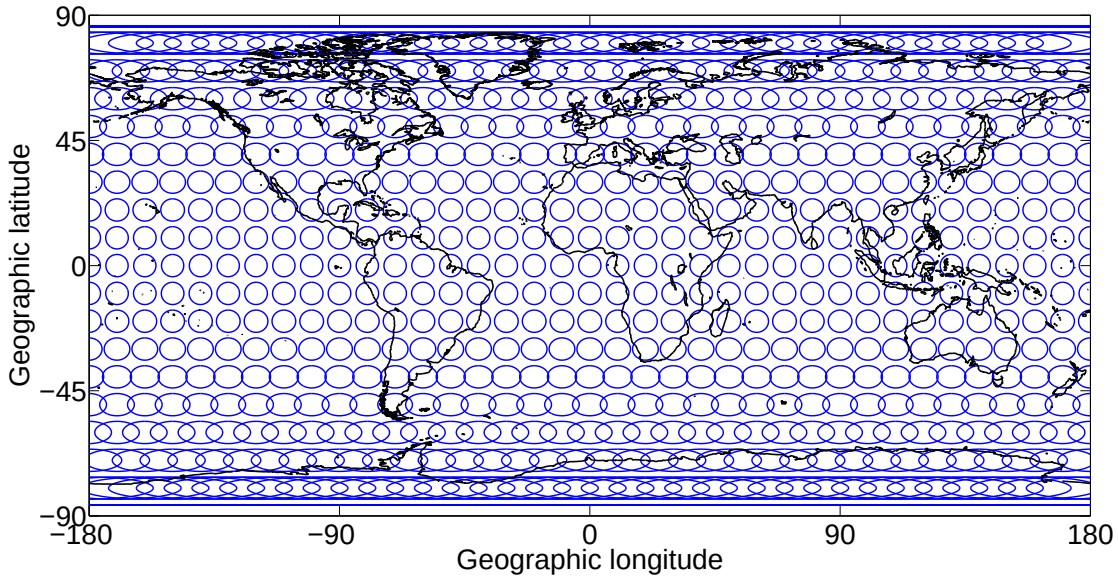

Figure 2.4: Schematic distribution of cylindrical VO bins, each of 400 km radius (roughly equivalent to an 8° width at the equator) and spaced evenly on a 10° latitude-longitude grid.

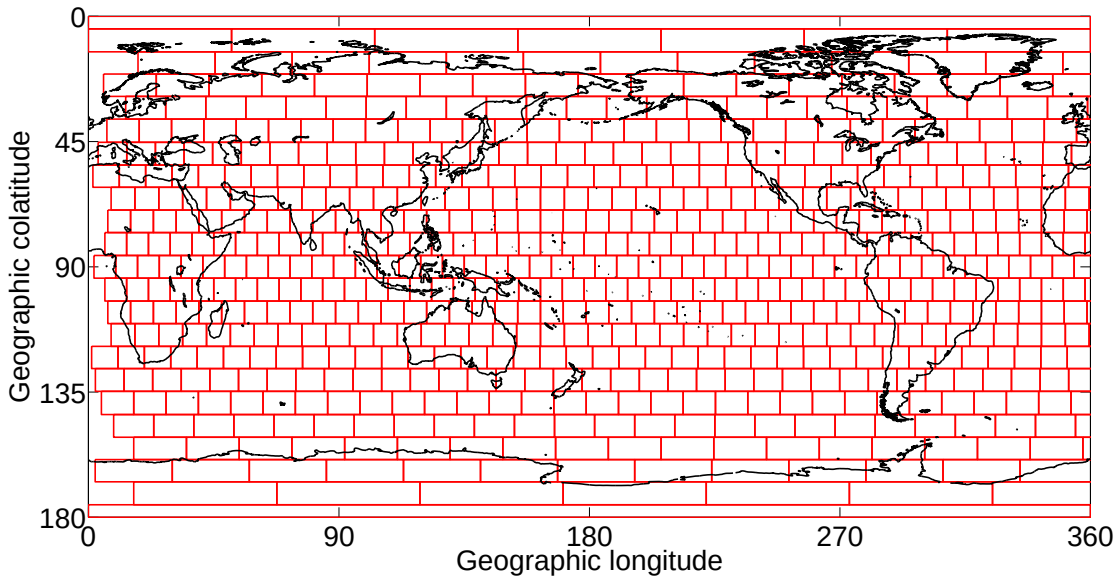

Figure 2.5: Distribution of 648 equal-area tesseral VO bins, each roughly 8° wide at the equator.

Beggan (2009) used both VO method variants separately to calculate monthly SV directly from the magnetic vector data of the CHAMP satellite (Reigber et al., 2002), for the time period 2001-2007. CHAMP was a near-polar orbiting satellite in LEO, launched at an initial altitude of 460 km, which precessed slowly in local time at a rate of 2.8 hours LT per month. It collected magnetic vector data at a rate of 1 Hz. Beggan's estimates of the SV were used to construct a series of monthly flow models of the liquid iron at the surface of the outer core. Assuming that the magnetic field is advected by the liquid iron, the resulting flow map can confer a more accurate prediction of SV than is possible with the (more arbitrary) approaches of linear or spline-fitted interpolation to extrapolate a future scenario of magnetic field morphology from existing values. Beggan found that the flow models exhibited rapidly varying patterns which changed in magnitude, direction and pattern from month to month. The variation of the patterns was too rapid to be physically indicative of sources internal to the Earth. The flow model was used to forward model the SV at the VO locations, which was then compared with the SV (*e.g.* equation (2.3)) inverted to make the flow model. Analysis of the global means of these residuals showed a significant correlation (0.66) with the first time derivative of the monthly mean Dst (Disturbance Storm-Time) index. The time series of Beggan's globally averaged flow model residuals and of the first time derivative of Dst's monthly mean are shown in Figure 2.6. The regularised flow model is based on a degree 14 SHA of SV data, and the residuals will have contributions from all degrees, so the correlation of the residuals with the Dst index (which parameterises a degree 1, order 0 field) was surprising. This contamination of the flow models with the low-degree (of spatial complexity) external fields was unexpected, and prompted further tests into its origin.

Beggan constructed a series of experiments designed to mitigate the contamination in the flow models and assess the source of the rapidly varying patterns. Correcting for magnetospheric fields via the CM4 (Sabaka et al., 2004) model reduced the correlation with the Dst index to 0.41, but did not significantly reduce the absolute magnitude of the residuals (Beggan (2009, page 120)). This indicates that there is a significant non-magnetospheric component to the contamination. Selecting only data from the nightside to construct the VOs, and again

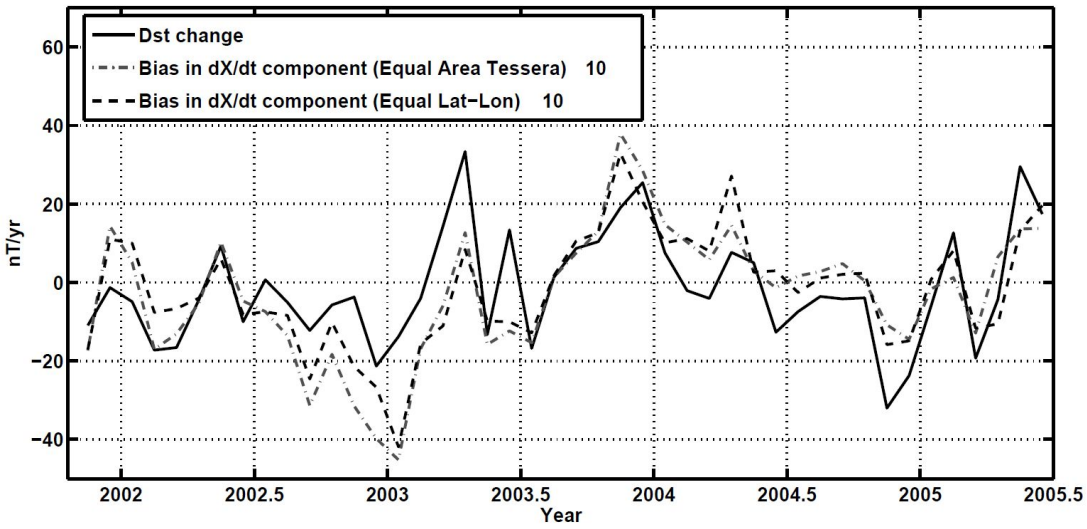

Figure 2.6: Globally averaged residuals to a flow model computed at spherical harmonic degree 14 from VO-derived SV estimates from CHAMP satellite data, compared against the first time derivative of the monthly mean Dst index. Chart is from Beggan (2009). The two dashed lines pertain to the different VO bin variants: 'Equal Lat-Lon' is shown in Figure 2.4, and here 'equal' refers to the 10° spacing between VOs. The 'Equal area tesseral' method is shown in Figure 2.5, and here 'equal area' refers to the equal area partitioning of the sphere. The correlation between the dashed and solid lines is 0.67 for the tesseral method and 0.66 for Lat-Lon.

correcting for magnetospheric fields with CM4 reduced the correlation with Dst to 0.10, but again, had a minimal effect on the absolute magnitude of the residuals. This indicates that the contamination is composed of at least magnetospheric and (predominantly dayside) ionospheric sources, but that it is not reduced appreciably by selection and correction for these signals, indicating an additional, unexplained source for the residuals. Additionally, utilising the CM4 magnetospheric field correction introduced extra patterns of sectorial banding into the geographic distribution of the residuals, implying that the model's correction is imperfect. Avoiding the CM4 correction and reducing the magnetospheric field signal by using only quiet-time data in constructing the VOs produces flow model residuals which are smaller and more random than those from all field activity levels, but which are still significant.

The fact that the absolute magnitudes of the residuals are independent of (rather, not lessened on average) application of the CM4 corrections, and that the Dst SV correlation also appears independent of the absolute magnitude of the residuals, is evidence that significant

parts of the contamination signal may not be described well by either Dst or CM4. Beggan (2009) concluded that the VO method is biased by both the UT and LT changes in the external fields. VOs do not make the external fields average to a zero-mean over the course of a month. Additionally, rather than averaging out geographically-diurnal oscillations, the VOs sample only a limited band of local time (around 2.8 hours for each month). This leads to problems when directly computing SV from monthly VOs, since the LT band sampled at t-6 months will not be the same as that sampled at t+6 months. This means that unless fields which are arranged in LT are accounted for properly in the calculation of the VO solutions, the resulting SV will contain apparently rapidly-varying signal related to the LT content of each month of VOs. Beggan (2009) showed using a synthetic data set that the sectorial banding patterns in the flow model residuals (linked to the rapidly varying flows resolved with this method) were due to a combination of the orbital drift of the satellite and the VO binning method used.

The correlation of the flow model residuals with Dst is indicative of aliasing of the low-degree external fields by the VO data binning method. However, the existence of this aliased signal in the residuals to the flow models is actually a secondary concern. What is of greater importance is whether this aliased signal is being introduced into the global models built from the VO solutions. The form of SH inversion used to construct the flow models is commonly employed in magnetic field research. Furthermore, the use of satellite data in studying geographically-fixed phenomena is an increasing trend. If the contamination affecting the VOs is indicative of the 'general case' in studies of the magnetic field which use satellite data, the impacts could be widespread. Attempting to better understand and mitigate the contamination in the VOs is therefore a valid concern, and is the focus of this chapter.

2.2.4 The *Swarm* satellite constellation

The European Space Agency (ESA)'s upcoming *Swarm* mission (described in Friis-Christensen et al. (2006)) is a constellation of three vector magnetometer-bearing satellites in near-polar LEO. The *Swarm* constellation geometry is unique. Two of the satellites (A and B) fly side-

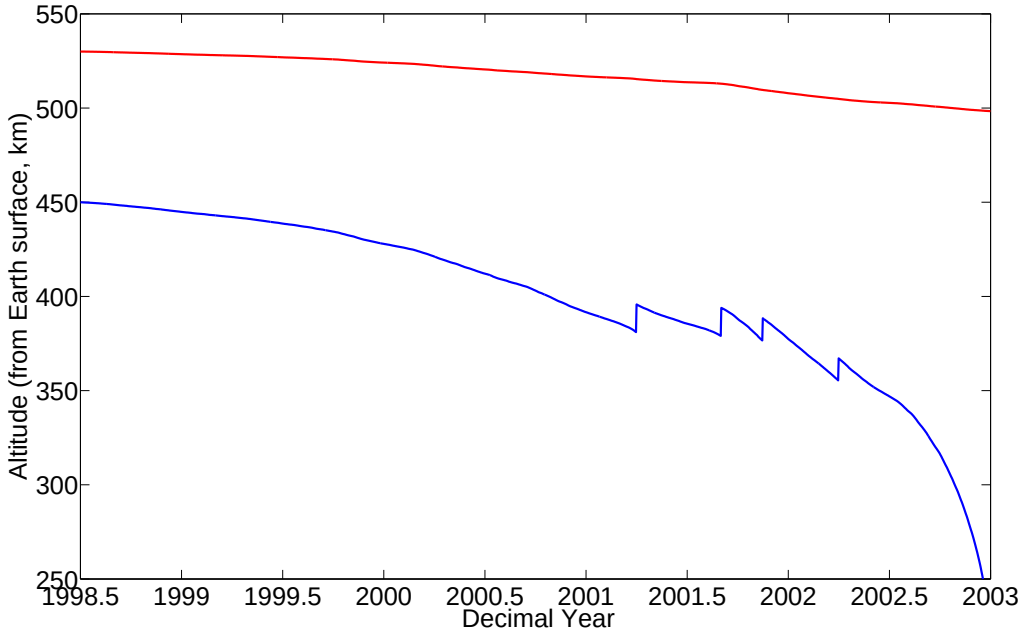

Figure 2.7: Modelled altitudes of the *Swarm* mission satellite A (blue) and C (red). The four altitude boosts applied to *Swarm* A appear as near-instantaneous 'jumps' in altitude. Reproduced from Friis-Christensen et al. (2006) using E2Eplus data (described later, in section 2.2.7), which is why the dates are in the past.

by-side with a separation of 150 km at an initial altitude of 450 km. The third satellite (C) is launched at an altitude of 530 km, initially in the same LT plane as A and B. The projected altitude decay of the satellites is shown in Figure 2.7. The two orbital planes of A and B, and C, are designed such that *Swarm* C precesses slightly slower than A and B in LT (Olsen et al., 2006a), as shown in Figure 2.8. At 3.5 years into the mission, the satellites' orbital planes are separated in LT by 6 hours, providing an improved simultaneous sampling of the external fields, and improving their representation in magnetic field models.

In this chapter, I use the unique constellation geometry of the *Swarm* mission to assess whether the parts of the VO contamination which are due to the LT arrangement of the external fields can be better mitigated, such that the contamination can be more fully reduced. In the process, an improved description of the contamination is expected, such that its effects in other uses of satellite data can be made clear. At the time of writing, *Swarm* is yet to launch so I use synthetic *Swarm* data from a sophisticated model designed to mimic the magnetic fields that

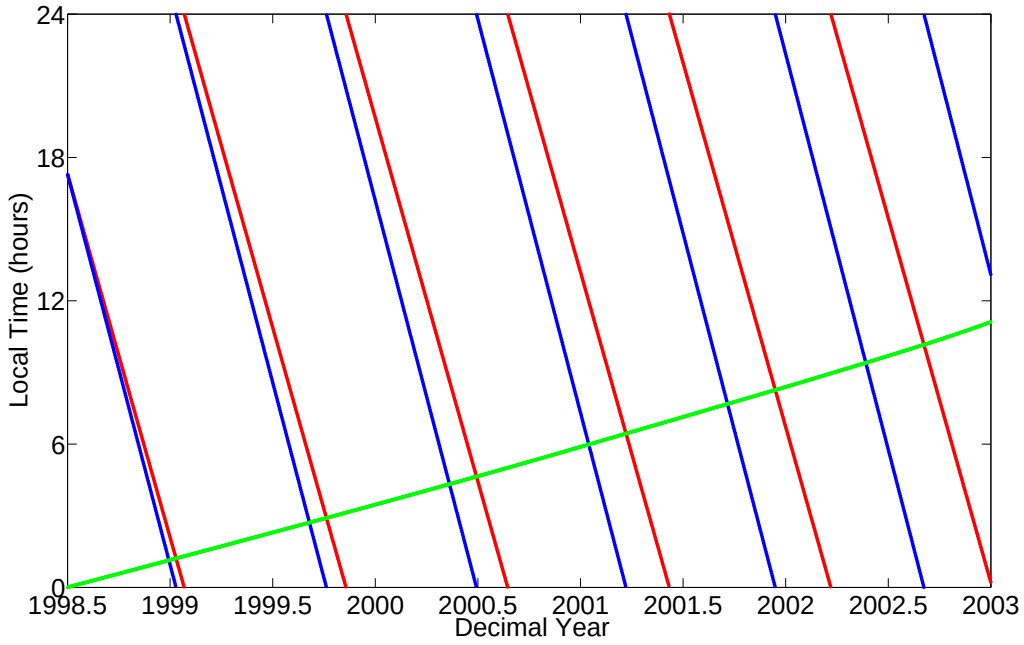

Figure 2.8: Modelled LT precession of the equatorial crossing of the ascending orbits of *Swarm* A (blue) and C (red). The green line is the difference between the two, indicating the divergence of the two orbital planes in LT. Note that a difference of greater than 6 hours corresponds to the two orbital planes beginning to converge. Reproduced from Friis-Christensen et al. (2006) using E2Eplus data (described later, in section 2.2.7), which is why the dates are in the past.

Swarm will sample. The model is the End-to-End (plus) *Swarm* mission simulator produced by Olsen et al. (2007), which is discussed in detail in a later section.

2.2.5 Summary of study objectives

It is desirable that the VO solutions are entirely free of contamination from external fields. However, it is important to recognise that the VOs were designed to represent the magnetic field in the same manner as the monthly means of permanent observatories. The permanent observatory monthly means may also be partially contaminated with the signals of external fields. Note that the original VO method eschewed data selection, preferring that the input data be the same as those used in the permanent observatory monthly means. Since the VOs appear not to sample the field in the same way, I am forced to recognise that the VOs themselves are a form of data selection, but one which is spatiotemporal, rather than temporal. I mention this in order to put the results of this chapter in proper context. The VOs likely represent the general issue of using LEO satellite data in analyses of geographically-fixed features, but they may

likewise respond to, or exacerbate, certain effects which studies not using VOs can neglect. It is expected that the efforts to reduce the contamination in the VOs may not be able to remove it completely. However, the identification of the causes of the contamination are intended to prove useful in future efforts, such that the user has a better control over which parts of the external fields the VOs sample. In this chapter I aim to characterise the VO contamination, from its starting level as a low-frequency set of external fields, to its aliased form in the VO solutions and the impact of these on a global SH model. Strategies for reducing the impact of the contamination on the VO solutions will be employed and their efficacy assessed. A major objective of this work will be to examine the improvements in external field representation offered by satellite constellation missions, *Swarm* in particular.

The main research questions of this chapter are as follows:

- Why do the VOs fail to sample the external fields such that they have a zero-mean over the course of a month?
- How much of the VO contamination is being aliased into the degree-14 SH model?
- Can the VO contamination be reduced significantly, and what is required to achieve this whilst achieving the same aim of calculating SV directly from the satellite data?
- What can the synthetic data VOs tell us about VOs applied to real data?
- Can *Swarm* constellation satellite magnetic data be used to estimate core field SV without incurring significant data loss through selection methods?

2.2.6 Likely sources of contamination in the VO solutions

Beggan's (2009) investigations into the source of the VO contamination showed that even when using only nightside data and correcting for magnetospheric fields, the flow model residuals still had significant magnitudes and bias. Although a combination of UT and LT effects is likely as the major cause of the contamination, it is worth assessing the assumptions made in producing a VO solution, in case a source of contamination is being overlooked. The simplifying assumptions required to produce a VO solution can be split into three broad categories:

temporal simplifications, spatial simplifications and (magnetic) source simplifications. I will introduce each in turn here.

The temporal simplifications involved in applying the VO method can be split into 'UT' and 'LT' effects. Firstly, it is assumed that the sampling of the UT-variant external fields is dense enough such that their representation in the VO is an appropriate 'reconstruction' of the original signal. In other words, for a VO to represent the external fields in the same manner as a ground-based observatory, the 'sampling frequency' of the VO must be at or above the Nyquist frequency (Weisstein, 2003) of the time series of the external fields, as measured on a geographically-fixed platform. This is a complex relationship between VO size, satellite orbit rate, Earth rotation rate, geomagnetic latitude and the number of satellites – I will discuss it in more detail in section 2.4. Secondly, it is assumed that the LT-content of each VO is identical to its peers in the same month. Put another way, the data in each VO are composed of measurements from two bands of LT, where each band is roughly 2.8 hours in width, and the bands are separated by 12 hours. If these two LT bands have significantly different magnetic environments (*e.g.* noon and midnight), then it is crucial that each VO for the month in question samples the same amount of each LT band. If it does not, then it may produce a different local solution to its neighbours, leading to spurious global structure. In short, the 'LT' effects describe the evenness of temporal data coverage throughout the month in a certain VO, in relation to neighbouring VOs. The 'UT' effects describe the typical temporal density of the sampling within a specific VO, in relation to the temporal behaviour of the external fields.

The VO method also requires that some spatial simplifications be adopted, such that a local solution can be obtained from a cloud of data 400 km in radius. Whilst a model of the static magnetic field is removed from each data point in the cloud, the static field prediction is added back on only at the centre of the VO. The centre of the VO is a fixed point, and is not the centroid of the data cloud. The static field varies non-linearly across the solution region, and any unevenness in the geographic distribution of the data cloud will incur a bias in the local solution. This bias will be amplified according to the magnitude of the static field, which can vary (albeit mostly linearly) by up to 4 nT per km at mid-latitudes (Mandea & Olsen, 2006). I

have run tests (not shown) which show that the spatial biases affecting the VOs are minimal, mostly because the VO bins are chosen to be equal-area. These tests involved predicting the CHAOS-2 main field model of Olsen et al. (2009) at the location of the VO centroid, and again at the centroid of the data locations within the VO (for a sample month). The differences in the two sets of predicted values were negligible and exhibited no coherent global pattern. If the VOs were not equal-area, a systematic bias might have occurred, though it would likely still be negligible. It is worth mentioning that the definition of the 'higher order terms' ignored by each VO is determined by the spatial size of the bin. I use the 648-bin tesseral VO approach of Beggan (2009) throughout this study, retaining an 8° (roughly 900 km at the equator) longitudinal diameter region for each VO. Langel & Hinze (1998) stated that [lithospheric] anomaly magnitudes at an altitude of between 400 and 600 km have amplitudes within the range ± 15 nT with wavelengths between 700 and 3000 km. The Nyquist theorem states that the smallest wavelength that can be restored is twice the maximum sampling gap. So, if 2*(maximum data separation in the VO) is greater than 700 km, the lithospheric fields will be under-sampled and a false signal will be created in the SV. The limit of 700 km will change as satellite altitude decreases, but it is unlikely that the lithospheric fields are being misrepresented in my analyses.

Lastly, the VO process involves simplifications of the magnetic field sources. I have already discussed Beggan's (2009) identification of the negative impacts associated with computing the SV from VOs which take different LT bands as input. I have also discussed the assumption of the VO method that the magnetic fields within the source region vary linearly. An additional simplification is the common step of ignoring the 'in-situ' magnetic fields resulting from electric currents at satellite altitude. The VO method entails a potential field approximation, so the VO solution should remove non-potential (*i.e.* in-situ) fields, making the local solution akin to the observatory monthly mean. However, the efficacy with which it does strip out these non-potential signals is not well known, and may represent an additional form of contamination affecting the solutions. All other things being equal, the effects of the non-potential fields should self-cancel when the SV is calculated. Unfortunately this cancelling effect does not take into account the tendency of the altitude decay rate of LEO satellites to increase with

time. The acceleration of the altitude decay presents another (small) contamination effect in the VOs. This is added to by the tendency of the satellite orbit's ellipticity to change with time (becoming more circular), though this is an even smaller effect.

From the results of Beggan (2009), it may come as no surprise that the temporal simplifications are by far the most damaging to the VOs. In the following sections I assess why this is the case, and how to mitigate the resulting contamination effects.

2.2.7 Synthetic constellation satellite data – the E2Eplus model

I aim to assess the efficacy of the *Swarm* satellite constellation geometry in mitigating the VO contamination. Since the contamination was identified by Beggan (2009) as stemming from fields external to the satellite orbit, the effectiveness of the E2Eplus model's parameterisation of the external fields is an important consideration. In this section I examine the make-up of the E2Eplus model, which is described in full in the reports of Olsen et al. (2006a) and Olsen et al. (2007). The E2Eplus model's core field parameterisation is based on the comprehensive model CM4 (Sabaka et al., 2004), with updated parameterisations of the external, lithospheric and induced fields, and a noise model appropriate to the *Swarm* mission, each of which is discussed in turn.

The CM4 model adopts the comprehensive approach to representing the geomagnetic field. This means that the positions of the various field sources are considered with respect to the locations of the sampling shells of both ground-based observatories and satellites. In this way, the contributions from the various field source regions are separable in a coestimation of all the model parameters. The E2Eplus model uses CM4 for the core (up to SH degree 13), lithospheric (degree 65), ionospheric and induced field predictions. The SV is described by cubic splines (a type of curve constraining temporal behaviour) given by CM4 up to degree 10, with degrees 11-19 appended from a separate model and assumed to be linearly variant in time. This is more than adequate for this study. A separate model was also used to increase the lithospheric model's resolution to SH degree 250. Ionospheric and related induced fields are

unchanged from CM4, in which diurnal frequencies are modulated linearly by the $F_{10.7}$ index of solar flux density. The largest change is in the magnetospheric field representation, which is replaced entirely from CM4's version. To produce the magnetospheric field model, ground-based observatory data were corrected with CM4's main field, secular variation and ionospheric (primary and induced) predictions, and subjected to a series of hour-by-hour SH analyses up to degree 3 and order 1. The induced fields were calculated from the resulting SH coefficients using a 3D model of mantle electrical conductivity. To compute the toroidal magnetic fields resulting from ionospheric F-region electric currents at satellite altitude, data from the Ørsted satellite were used to compute a toroidal field representation at 750 km altitude, which is then considered altitude-invariant. This is imperfect, but better than assuming no toroidal fields. An investigation of the F-region currents forms the basis of a later chapter of this thesis. Finally, noise is added that is correlated in time, but uncorrelated among the components, and scaled to the projected performance (in nT-accuracy terms) of *Swarm*. The E2Eplus model is no longer based on the comprehensive approach, but is now a set of serial modular parameterisations, which nevertheless entail a more sophisticated representation of the external fields than is provided by CM4.

The first stage in producing data predictions from the E2Eplus model is to synthesize the *Swarm* satellite orbit predictions. Two synthetic 'spacecraft' with initial altitudes of 450 km and a third with initial altitude of 530 km are assigned approximately 7 km/sec velocities. The equations of motion are solved on a 1-minute basis, with atmospheric drag provided by the MSIS-86 model (Hedin, 1987) and drag coefficients tailored according to experience from the CHAMP mission. Four in-orbit altitude manoeuvres (two of 12 km, two of 15 km) sustain the lower satellites' altitudes for the required mission length of 4.5 years, prior to 're-entry' of the lower satellites at an altitude of 200 km (refer to Figure 2.7). Olsen et al.'s (2007) comparison of the CHAMP altitude decay with that of the synthetic orbits shows good agreement, indicating a realistic trajectory for the synthetic *Swarm* spacecraft.

The E2Eplus model is computed as static hourly expansions for the epochs from 00:00

on the 1st July 1998 to 00:00 on the 1st January 2003, encompassing the 4.5-year time span of the synthetic orbits. The synthetic launch time was chosen to be a full solar cycle before the planned launch in 2010, though the *Swarm* mission has since experienced delays in the actual launch date. Forward modelling of the data is performed via interpolation of the hourly coefficients to a 15-second sampling interval, followed by data predictions of the core, SV, lithospheric, magnetospheric and induced, ionospheric and induced, toroidal and noise contributions at each of the three synthetic satellite locations[2]. The sampling rate of the synthetic *Swarm* constellation is therefore $0.0\dot{6}$ Hz, in comparison to CHAMP's 1 Hz. This reduction in data density does not impair the ability to produce VO solutions, nor does it reduce the appropriateness of the synthetic data as a real-world analogue. However, the model is certainly not a uniformly-good representation of all parts of the magnetic field, which I will discuss later in this chapter.

The ECEF (Earth-Centred, Earth-Fixed) Cartesian coordinate system is defined such that the x-axis points from the centre of the Earth towards the Greenwich meridian in the geographic equatorial plane. The z-axis points towards geographic North, and the y-axis completes the right-hand triad (*e.g.* Hapgood, 1992; Langel & Hinze, 1998). Spherical ECEF coordinates (r, θ, ϕ) are defined such that θ (co-latitude) is the angle from the z-axis and ϕ the angle from the x-axis in the x-y-plane, and positive towards East. The E2Eplus data are perfectly aligned in the spherical polar ECEF frame, whereas the CHAMP data are subject to attitude errors resulting from thermally-driven misalignment of the optical bench, which is the part of the satellite which houses the star cameras and vector magnetometer, shown in Figure 1.11 (Olsen et al., 2006b, page 70). This difference in alignment accuracy does not appear to affect my analyses, but checking how similar the E2Eplus VO solutions are to those from CHAMP data will constitute the initial part of the results section.

[2]E2Eplus synthetic data were obtained from `ftp://ftp.spacecenter.dk/data/magnetic-satellites/Swarm/E2Eplus/constellation_4/15secs_crust-250/` on 6th October 2009.

2.2.8 Reference frames

Satellites in LEO fly through the edges of a highly temporally variant system of external fields, and sample a highly spatially variant system of static magnetic fields. Discussion of the causes of the VO contamination with respect to the magnetic field sampled by permanent observatories will involve a number of reference frames, which I summarise here to avoid confusion.

Temporal reference frames:

- UT frame: VO solutions or SHA model in linear progression of time.
- SV frame: VO solutions or SHA model, after the first difference is taken. Static fields cancel out.

Spatial reference frames:

- GEO frame: UT or SV frame, in GEO coordinates. Spatial differences are due to geographic distribution.
- LT frame: coordinates are geographic colatitude and UT hours difference in longitude from that of the sub-solar point, divided by 360°. Spatial differences are used to define which current systems are dominant in the local magnetic environment. Discussion of the LT-morphology of the external fields thus refers to their variation in LT, which can be interpreted as spatial, rather than temporal, changes in the LT frame.

Note that when I refer to the UT-variance of the external fields, a GEO coordinate system is assumed unless the current system is zonally invariant in LT. In other words, diurnal variation is assumed present in all discussion, according to Earth (hence, VO) rotation.

2.3 Results

2.3.1 Replicating the contamination in synthetic data

The first stage of the VO study was to compute the residuals to a global SH model made with synthetic-data VOs, to see if the same contamination exists in the synthetic data as it does in the real data. The VO setup I use is the same 648-bin equal-area tesseral partitioning of the

sphere as that shown in Figure 2.5[3]. In all instances, the polar bin solutions are rejected due to their high errors. Note that a time series of the difference in the first Gauss coefficients $\left(g_1^0\right)$ of global degree 14 SHAs of the tesseral and cylindrical VO variants using E2Eplus data (not shown) exhibits a minimal difference with a 4.5-monthly oscillation (corresponding to the period of the LT coverage of the satellite). This oscillation is likely due to the duplication of ionospheric signal at midlatitudes in the VO cylindrical arrangement, in which the bins overlap as the poles are approached. The cylindrical method is not discussed further. The model used to provide $\mathbf{B}_{MF}$ estimates is CHAOS-2s[4], produced by Olsen et al. (2009). Monthly VO solutions were produced for all 54 months of synthetic E2Eplus data for *Swarm* satellite A, resulting in 42 monthly sets of SV.

The result of Beggan (2009) in which contamination was identified in the VOs pertained mostly to the residuals to core-mantle boundary flow models derived from an inversion of VO SV data. I do not compute outer core flow models in the same manner as Beggan (2009), but still require an analogous global smooth model which brings all the separate VO solutions together into a single function, so that the contamination in the synthetic data can be assessed in a similar manner to the real data (I only treat synthetic data in this study). As well as assessing the contamination in the VO solutions, I plan to assess the amount of contamination signal aliased into the global model, so the amount of smoothing applied in the process of computing the SHA is important. Here, some of the specific aspects of the SHA I use are summarised.

The SHA I use is an internal-assumption (the first part of the right-hand side of equation 1.1), degree-14 expansion at a reference altitude of 400 km. The problem to solve is

$$\mathbf{d} = \mathbf{G}\mathbf{m}, \tag{2.4}$$

[3]Matlab code to compute the partitioning of the sphere was supplied to me by Ciaran Beggan, 2009, originally obtained from Paul Leopardi.

[4]using the spline coefficient file available at `http://www.spacecenter.dk/files/magnetic-models/CHAOS-2/CHAOS-2s_alpha_04.mat`, but provided by Ciaran Beggan.

where **d** are the data I wish to represent via the SHA (the VO solutions of SV for a given month, less the polar bins), **G** is the equations of condition matrix (Menke, 1989) and **m** is the model composed of Gauss coefficients, which my estimate of is $\hat{\mathbf{m}}$, the regularised (smoothed) L2 norm least-squares solution, given by

$$\hat{\mathbf{m}} = \left(\mathbf{G}^{\mathsf{T}} C_e^{-1} \mathbf{G} + \lambda C_m^{-1}\right)^{-1} \mathbf{G}^{\mathsf{T}} C_e^{-1} \mathbf{d}, \tag{2.5}$$

where C_e^{-1} is the inverse of a matrix of the covariances of the original data, C_m is the model covariance matrix through which regularisation is applied, with λ the damping parameter. C_e is assumed diagonal (but will have non-zero off-diagonal elements due to the first-difference applied to the data), and is used to normalise the data by their standard deviations, helping prevent a poorly-conditioned normal equations matrix $\left(\mathbf{G}^{\mathsf{T}}\mathbf{G}\right)$. The errors on the data are estimated in this case from the misfit of the final weighted VO model to the data in the IRLS iteration. This is not an ideal representation of the errors of the original (pre-VO) data as the VO model is weighted, but it will certainly suffice for my purposes. The *a priori* constraint λC_m^{-1} applies damping to smooth the model solution according to a minimisation of the square of the radial magnetic field integrated over the core-mantle boundary (*e.g.* Parker & Shure, 1982), and C_m^{-1} is constructed via (Gubbins, 1983)

$$C_m^{-1} = \left(\sum_n \left(\frac{a}{c}\right)^{2n+4} \frac{(n+1)^2}{2n+1}\right) \mathbf{I} \tag{2.6}$$

where a is the reference radius of the SHA (6371.2 km), c is the core-mantle-boundary radius (3485 km), n is spherical harmonic degree and **I** is the identity matrix. As stated by Gubbins & Herrero-Bervera (2007, page 123), this general form of regularisation is tailored to satisfy the requirement that the field we solve for originates in the outer core – the regularisation increases with each spherical harmonic degree, producing a very smooth solution at the VO altitude. The precise damping is controlled by the constant λ, which can be selected by plotting

a 'trade-off curve' between the solution norm

$$\sqrt{\hat{\mathbf{m}}^{\mathsf{T}} \lambda C_m \hat{\mathbf{m}}}, \tag{2.7}$$

and the weighted residual norm

$$\sqrt{\left(\mathbf{d}-\hat{\mathbf{d}}\right)^{\mathsf{T}} C_e^{-1} \left(\mathbf{d}-\hat{\mathbf{d}}\right)}, \tag{2.8}$$

where $\hat{\mathbf{d}}$ is the SHA data prediction, given by

$$\hat{\mathbf{d}} = \mathbf{G}\hat{\mathbf{m}}. \tag{2.9}$$

The trade-off curve is shown in Figure 2.9. This curve is used to find the damping parameter giving a solution norm which is appropriately low, but which does not drastically increase the misfit to the data. This is typically found near the 'knee' of the curve. I use a damping parameter of 1×10^{-8} for all SHAs in this chapter.

The residual of the SHA data prediction to the VO solutions (the 'SHA residual') is given by

$$\mathbf{e} = \mathbf{d} - \hat{\mathbf{d}}. \tag{2.10}$$

Note that the use of an L2 norm is perhaps not best suited to the global distribution of the VO solutions (or resulting SV estimates), since there is good reason to suspect that they are not representative of a Gaussian distribution. However, the SHA is not required for the majority of the assessment of the contamination causes, as discussed in section 2.4.

Figure 2.10 (red line) shows the global means of residuals of monthly degree-14 SHAs to

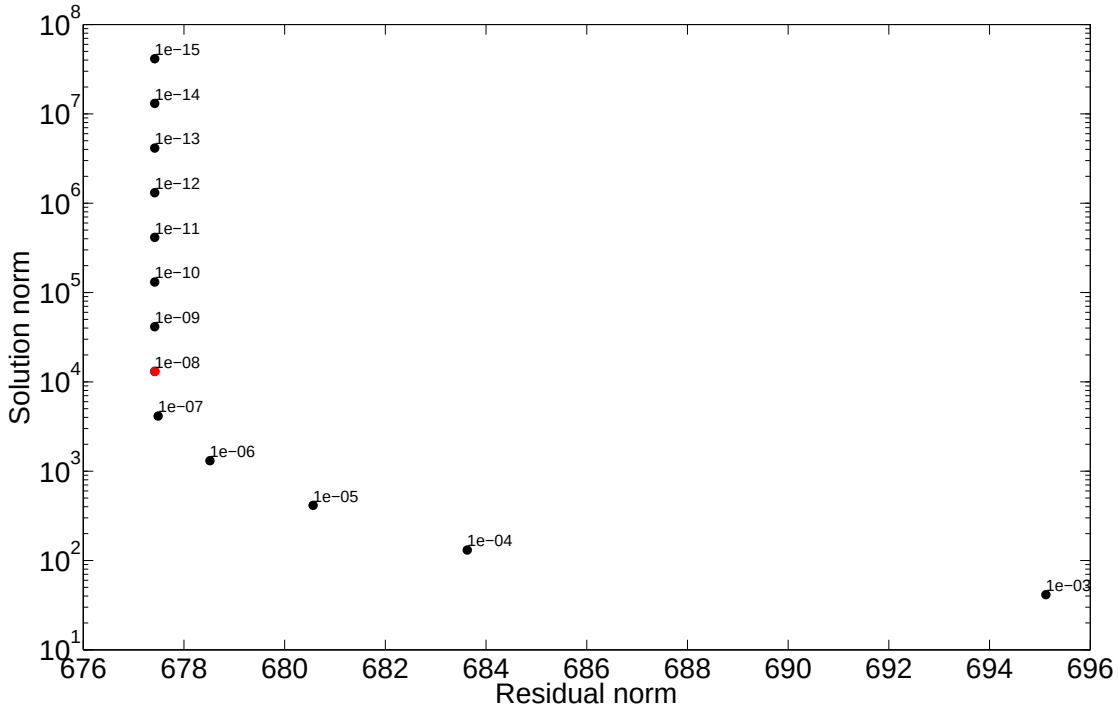

Figure 2.9: Trade-off curve for SH models of SV, computed from VOs of *Swarm* A synthetic data from July 1998 and July 1999. The trade-off values of solution and (weighted) residual norms are shown for 13 different damping parameters, the values of which are shown in text adjacent to each data point in the chart. The damping parameter used in this study is 1×10^{-8}, chosen as it lies near the 'knee' of the trade-off curve. The associated trade-off point is shown in red.

SV determined directly from the VO solutions of *Swarm* A synthetic data. The black line is the 12-monthly first time difference of the monthly mean Dst index ('Dst SV'). The Pearson's correlation (Press, 1992) of these two time series is 0.85. The green line is for *Swarm* C data, treated in the same manner as *Swarm* A data. The *Swarm* C residuals have a correlation of 0.79 with the Dst SV. The SHA residuals for *Swarm* B (not shown) are effectively identical to those from *Swarm* A: I will not consider the data from *Swarm* B further in this study.

Comparison of Figures 2.6 and 2.10 shows that despite different input data and a different global model, a very similar contamination is present in the synthetic data as in the real data. This indicates that the causes of the contamination affecting the VOs are persistent, and that I can (provisionally) trust that the E2Eplus data are complex enough for the purposes of investigating the VO method. The contamination in the synthetic data is not identical to that

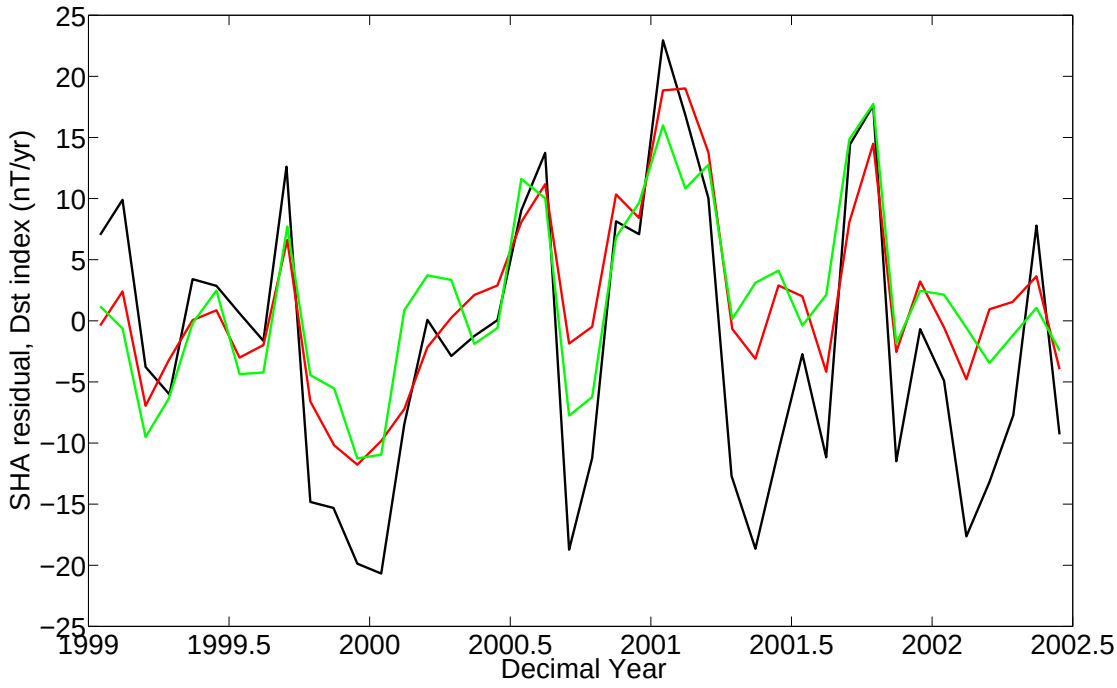

Figure 2.10: Global means of residuals to monthly degree-14 SHAs of SV determined directly from VOs of *Swarm* A (red line) synthetic data. The green line is the same process, using *Swarm* C data. The black line is the 12-monthly first time difference of the monthly mean Dst index.

in the real data, exhibiting a typically higher degree of correlation with Dst. This increased correlation might indicate a simpler LT-distribution of the external fields in the synthetic data. However, the correlation with Dst is different enough between *Swarm* A and C's SHA residuals to indicate that there is something other than just the symmetric ring current causing the contamination. Note that since the toroidal fields are altitude-invariant in the E2Eplus model, it now appears unlikely that altitude change throughout the mission is a significant cause of the contamination, though it may still present a small signal in real data VO solutions.

I have shown that the contamination in the synthetic data is of similar magnitude to that in the real data, and it is known from the results of Beggan (2009) that these magnitudes are damaging to studies of the SV. Since the amount of contamination introduced into the global SHA model is of interest to us, in Figure 2.11 I show SV time series from a reference model, and from a SHA model made from VOs. The black line is the global mean θ-component SV from the CHAOS-2 model predicted monthly at the VO locations, with a linear trend removed

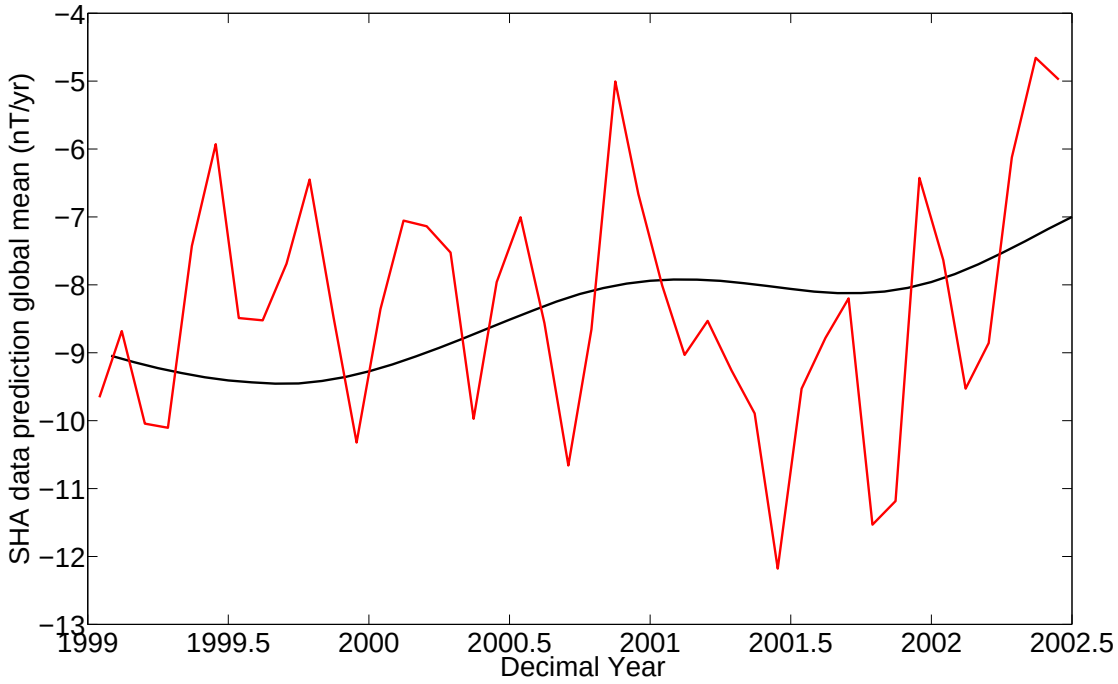

Figure 2.11: θ-component SV derived from global means of data predictions at 648 VO locations of monthly degree-14 SHAs of θ-component SV determined directly from VOs of *Swarm* A synthetic data (red line), and of CHAOS-2 (black line). Both time series have had a linear trend removed, in order to exhibit small temporal-scale variations.

in order to exhibit its small-scale temporal variation. The red line is the SV computed from the data prediction of a degree-14 SHA made from *Swarm* A VOs, again with a linear trend removed. This chart provides a 'benchmark' for the magnitude of contamination in the SHA model – any contamination effects in the SHA model global average data prediction on the order of this magnitude are detrimental to SV studies.

2.3.2 Ability of the constellation to mitigate the contamination

The contamination has been shown to exist in the synthetic data, so I now construct a set of VOs using all the data from all three satellites in the *Swarm* constellation in a single set of VOs. Each VO will then contain contributions from four bands of LT, rather than two. It is expected that as the *Swarm* constellation's orbits precess to a 90° relative angle, corresponding to a 6-hour LT difference, the four bands of LT will provide an increasingly-good representation of the LT-arranged external fields in the VO solutions. I expect the effects of this improved

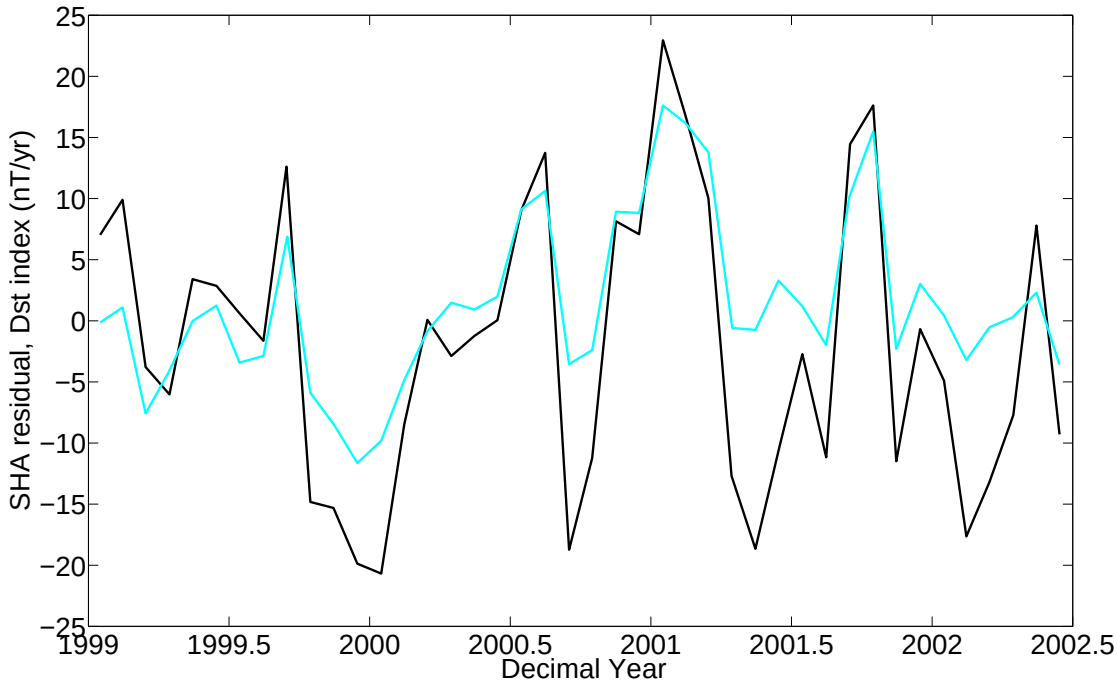

Figure 2.12: Global means of residuals to monthly degree-14 SHAs of SV determined directly from VOs of synthetic data from all three *Swarm* satellites (cyan line). The black line is the 12-monthly first time derivative of the monthly mean Dst index.

representation to be twofold: firstly, the contamination in the SHA residuals should change as the (more effectively sampled) external fields should be more easily separable from the internal fields by the internal-assumption SHA, and secondly, the external fields will present a more consistent 'spatial' signal in the LT frame, making their contributions more likely to cancel out in the SV frame (*i.e.* the external fields should behave more like static fields when sampled with the constellation). I therefore assess the effect of utilising the constellation geometry in the SV frame.

Figure 2.12 shows Dst SV (black line) with the global means of residuals to monthly degree-14 SHAs of SV determined directly from the VO solutions of synthetic data from all *Swarm* satellites (cyan line). The Pearson's correlation of these two time series is 0.85, indicating a minimal change in the contamination from the VOs using data from only a single satellite. Indeed, visual examination of the SHA residuals shows that even at 3.5 years into the mission when the *Swarm* constellation geometry is optimal for external field sampling, the contami-

nation is not significantly changed. This is a surprising result. The Dst index is corrected for ionospheric field variations (Sq), and that any ionospheric signal (which is predominantly arranged in the LT frame) will present a significant signal in the VOs which should be different from Dst. Furthermore, the *Swarm* constellation geometry has been validated in several detailed studies (*e.g.* Olsen et al., 2006a, 2007), and is known to be effective in representing the external fields. Yet my use of the same constellation geometry appears here to be having an insignificant effect in reducing (or even changing) the contamination. I am forced to recognise that since the E2Eplus model was made using an expansion of real data, the model itself may be contaminated. The separation of internal and external fields does not appear to be aided by the addition of more sampling locations in this case. It is important that the VOs themselves are assessed in detail, such that I can determine fully the cause of the contamination, and infer if it is having more widespread effects than just in the VO method.

In addition to assessing the effect of applying the *Swarm* constellation via the SHA residuals, I checked the effect on the VO solutions, the SV obtained directly from the VOs, the first Gauss coefficients of the SHA model, and the global mean data prediction of that model. All showed contamination reductions which were similarly only as small as those in Figure 2.12. To assess why the constellation geometry of *Swarm* is ineffective here, I will study each stage of the process involved in producing the SHA model, and will do this in the UT frame rather than the SV frame, in order to preserve the details of the original VO solutions and specifically to study the external field effects in more detail.

2.3.3 The contamination at each stage of the VO process, in the UT frame

The VO method is designed for determining SV. However, in my attempts to establish generic rules about the behaviour of the external fields, the first time derivative becomes a hindrance. The results in the remainder of this chapter will be in the UT frame, rather than the SV frame. Since I intend to study the external field's representation in the VOs, the SV signal is removed from the input data and the VO solutions are displayed without the main field model added back on – VO solutions and associated input data shown from this point onwards only have

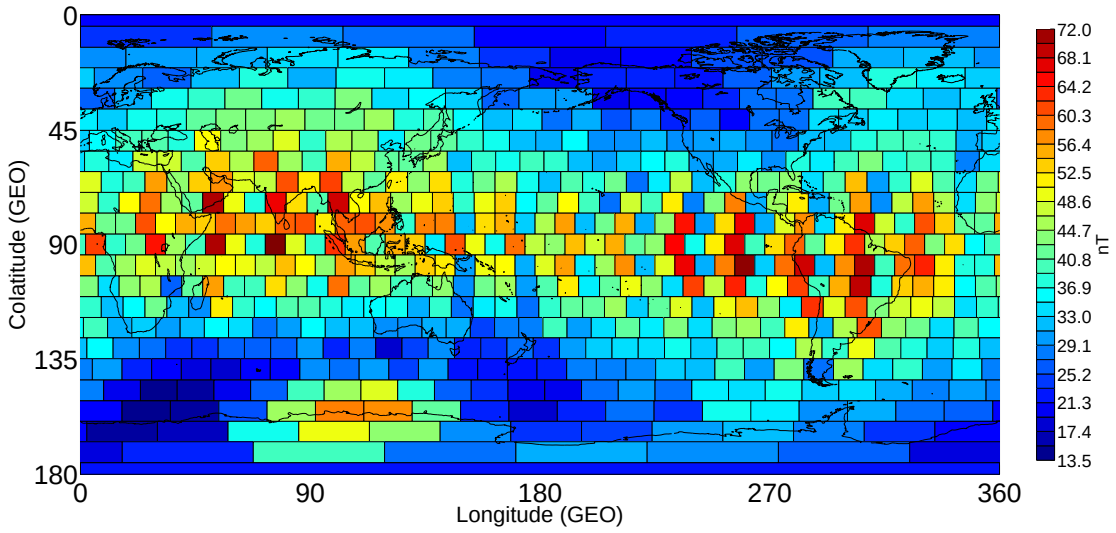

Figure 2.13: RMS values of the θ-component of the magnetic field from *Swarm* A data, October 2001. The RMS value is computed from the data in each VO bin after removal of main field and SV terms, in order to give a representation of the global distribution of data values for the month in question.

contributions from external and static fields.

There are three main 'stages' to the process I have discussed so far. The input data are first allocated to the VO bins. Second, VO local solutions are produced. Thirdly, a SHA produces a global model of these solutions. Here I summarise the representation of the contamination at each stage. Figure 2.13 shows the global distribution of 648 RMS (root-mean-square) values of the θ-component of the magnetic field (in each of the 648 VOs) for *Swarm* A, October 2001. Figure 2.14 shows the same, for *Swarm* C. These two figures give an indication of the global distribution of data values for 2001, October, as sampled by the VOs. Sectorial banding is seen in the sampled data from both *Swarm* A and C, though is much stronger in the latter. This difference is not due to sampling altitude – I will later show it to be due to how the data are sampled. Patterning similar to that in Figures 2.13 and 2.14 is seen in each month of VO data, regardless of external field activity level, but with varying magnitudes. This is the typical form the contamination of the data takes. I will discuss its causes in section 2.3.4.

Figures 2.15 and 2.16 show the sectorial banding contamination in the θ-component VO

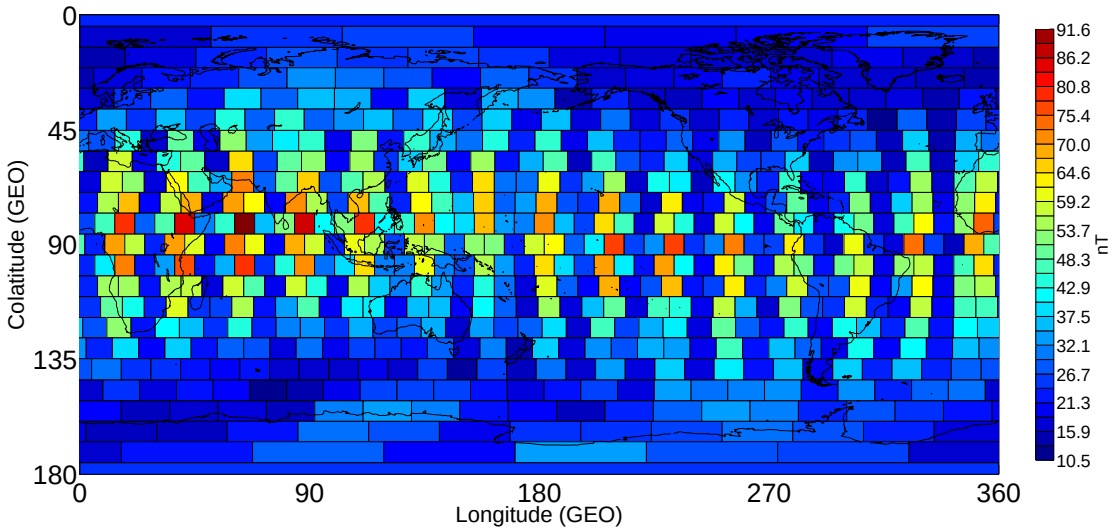

Figure 2.14: RMS values of the θ-component of the magnetic field from *Swarm* C data, October 2001. The RMS value is computed from the data in each VO bin after removal of main field and SV terms, in order to give a representation of the global distribution of data values for the month in question.

local solutions. Again, strong sectorial banding is seen, but it is reduced in severity from the input data (in terms of morphology rather than magnitude), particularly in the *Swarm* A VOs. The spatial pattern of the contamination in the UT frame varies between different months to an extent that it is nearly incoherent in time. If the annual difference (not shown) of the contamination signal is taken, the amplitudes of the contamination in the SV frame (in units of nT/yr) would have a nT range similar in magnitude to the contamination in the UT frame (shown in Figures 2.15 and 2.16 for the VO solutions). Therefore, because the contamination is temporally incoherent and rapidly variant, I assume that the magnitude of the contamination in the UT frame is simply related to the (typical) magnitude of the contamination in the SV frame. Hence I am able to continue studying the contamination in the UT frame whilst implying consequences for the estimation of SV from VO solutions. A comparison of Figures 2.15 and 2.16 with Figure 2.2 highlights that the contamination is comparable (in the sense stated above) to the magnitudes of the SV signal, presenting a significant bias. The misfit of the VO solutions to the input data (the 'VO residuals') for *Swarm* A and C in 2001, October are shown in Figures 2.17 and 2.18, respectively. The calculation of the residual of the VO solutions to the input data produces a multi-point distribution inside each VO, so the values shown here are their

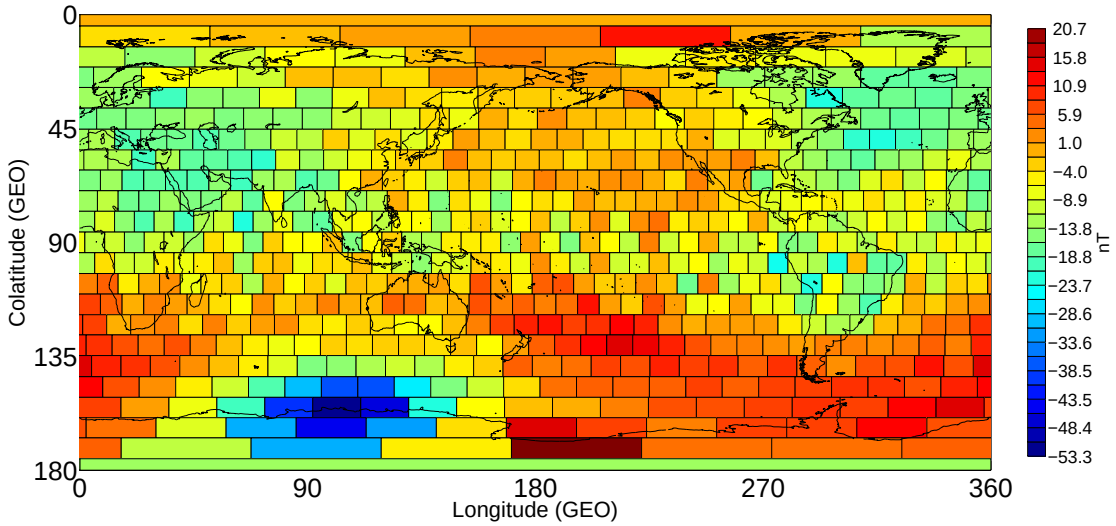

Figure 2.15: VO solutions for the θ-component of the magnetic field from *Swarm* A data, October 2001. Main field and SV are not included, to highlight the effects of the external fields.

RMS values. The magnitudes and patterns of the VO residuals are very similar to Figures 2.13 and 2.14, particularly for *Swarm* A. This similarity indicates that the IRLS solution process is rejecting significant amounts of the contamination present in the input data, but not enough to prevent it contaminating the VO solutions.

Figure 2.19 shows the data prediction of the θ-component magnetic field values for only external fields for *Swarm* C, October 2001. To obtain this distribution, VO solutions were computed from simulated data containing only the core and lithospheric fields, and separately, VO solutions were obtained from simulated data containing the core, lithospheric and external fields. Each of these VO solution sets was used to compute a degree 14 SHA model, and the data predictions (at the VO locations) of these two models were differenced. The data are then interpolated onto a 1° cell size mesh and a smooth surface fitted, with contours shown. Since the contamination in Figure 2.16 is spatially complex, it is mostly stripped out of the SHA model solution and ends up in the UT frame SHA residuals (not shown). However, identifiable parts of the contamination are indeed getting into the global model. I will show how much each external field source region contributes to the contamination of the SHA in section 2.3.5.

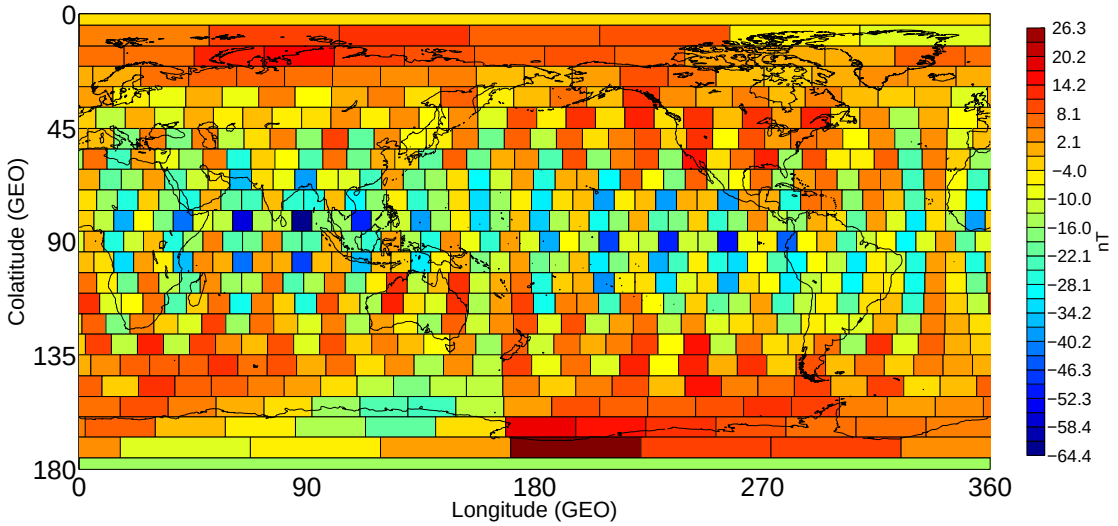

Figure 2.16: VO solutions for the θ-component of the magnetic field from *Swarm* C data, October 2001. Main field and SV are not included, to highlight the effects of the external fields.

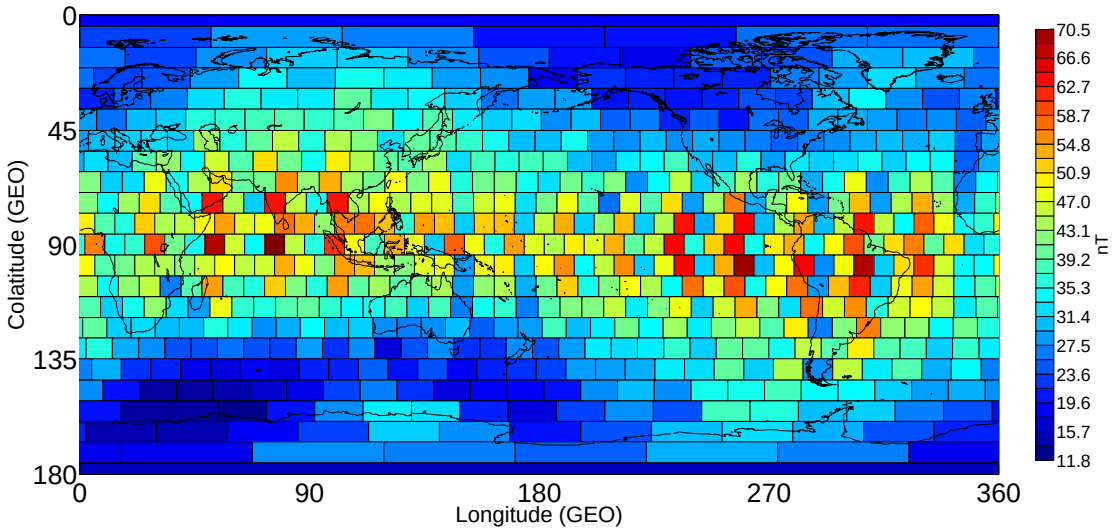

Figure 2.17: RMS values of residuals of VO solutions to the θ-component of the magnetic field from *Swarm* A data, October 2001. Main field and SV are not included, to highlight the effects of the external fields.

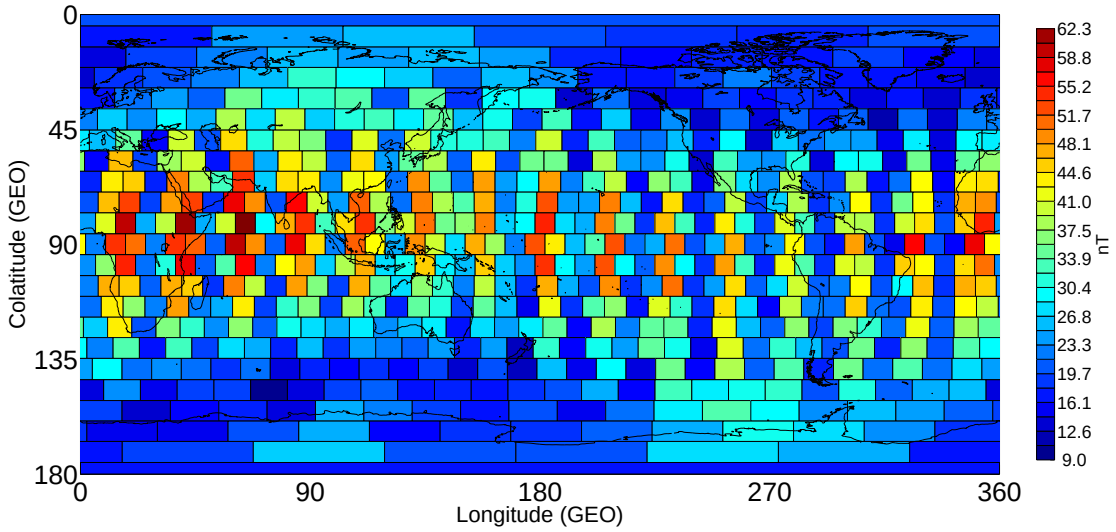

Figure 2.18: RMS values of residuals of VO solutions to the θ-component of the magnetic field from *Swarm* C data, October 2001. Main field and SV are not included, to highlight the effects of the external fields.

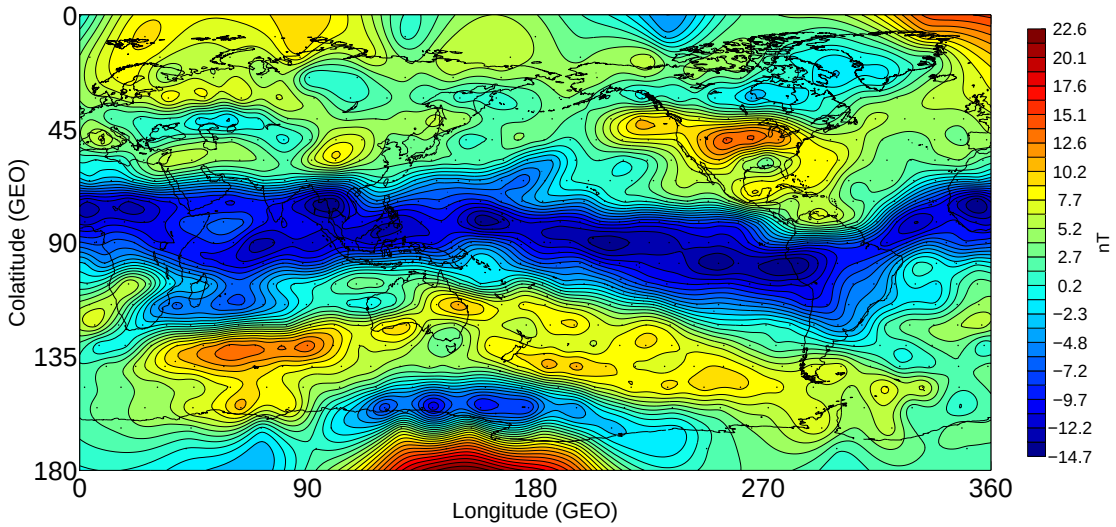

Figure 2.19: θ-component data prediction from a SHA model of VO solutions of the θ-component of the magnetic field from *Swarm* C data, October 2001. Only includes contributions from external fields. Further description is in the main text.

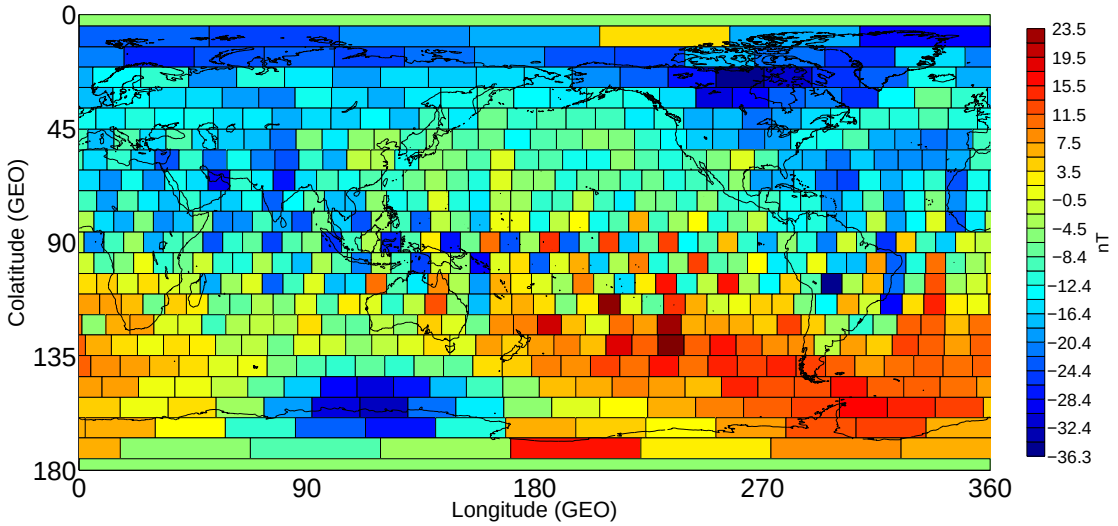

Figure 2.20: VO solutions for the θ-component of the magnetic field with *Swarm* A data, September 2000. Main field and SV are not included, to highlight the effects of the external fields.

2.3.4 Examining the causes of the contamination-related patterning in the VOs

The contamination at each stage of the VO process consists mainly of high-order sectorial banding, apparent in the VO-sampled magnetic data, and in the VO solutions. The SHA model contains a smoothed version of these contamination signals. The sectorial banding is the dominant form of the contamination, but infrequently a separate high spatial frequency signal is seen in the VO solutions, which has a (quasi-square) 'tiling' pattern. Figure 2.20 shows an example of this tiling pattern contamination in the VO solutions for θ-component data from *Swarm* A for September 2000. The two global representations of the contamination – the sectorial banding, and the tiling pattern – have different causes, and I will assess each in turn here.

The sectorial banding (so far shown most clearly in Figures 2.13, 2.14, 2.16, 2.17 and 2.18) is typically order-15 (*i.e.* 15 wavelengths in the equatorial circumference). In July 1998, the longitudinal displacement at the equator between successive orbits of *Swarm* A was 23.45°, indicating that every 15.35 orbits, it will sample the same geographic location again. This frequency changes throughout the *Swarm* mission due to altitude decay, but does not alter substantially. It is the orbit repeat rate which causes strong magnetic signal to appear in the order-15 pattern show in in above figures. The spatial complexity of the resulting contamina-

tion signal is part of the reason the sectorial banding does not appear in the degree 14 smoothed SHA model.

Maus et al. (2006a) also encountered sectorially-banded errors at multiples of order 15 in their *Swarm* synthetic data sets. The authors proposed that the periods of the *Swarm* mission when the satellites enter exact repeat orbits were the cause for these errors. Since the starting repeat-orbit frequency of *Swarm* A is non-integer, the first month of sampling does not conform to an exact repeat orbit scenario. However, recall that the VOs 'condense' spatial variations in the data into a single invariant solution location. Even if a satellite is not in an exact repeat orbit situation, the magnetic signal it records may still appear in the same set of VOs, repeatedly and at order 15 (an effective repeat orbit situation). *Swarm* C has a near-identical orbital drift rate to A yet frequently displays different magnitudes of sectorial banding, suggesting that the manner in which the VOs sample the satellite data is not straightforward.

To show how the *Swarm* satellites A and C can sample the same signal and produce different VO distributions, I sampled Dst index values to the UTs of the data inside each VO for October 2001, for *Swarm* A and C separately. Figure 2.21 shows the hourly Dst index values for October 2001. Figures 2.22 and 2.23 show the global distribution of the RMS of the sampled Dst values in each VO bin for *Swarm* A and C, respectively. Note that the same data (Dst) produces RMS distributions which are not only different to each other for the two satellites, but which are each similar to their respective counterparts in the VO solutions of E2Eplus data (Figures 2.15 and 2.16). The cause of the sectorial banding can be shown via the temporal distribution of data inside the VO, but this does not tell us anything that cannot be inferred from the VO distributions already shown. The sectorial banding results from the broad orbital-repeat rate of the *Swarm* satellites, the fragmentation of the satellite time series by the VOs, and the fact that the dominant external fields are characterised by rare large deviations rather than frequent smaller deviations (as discussed by Chisham & Freeman (2010) for the ionosphere, and attributable to the magnetosphere). The sectorial banding is the spatial manifestation of the 'UT effects' referred to in section 2.2.6 – I henceforth refer to the tendency of

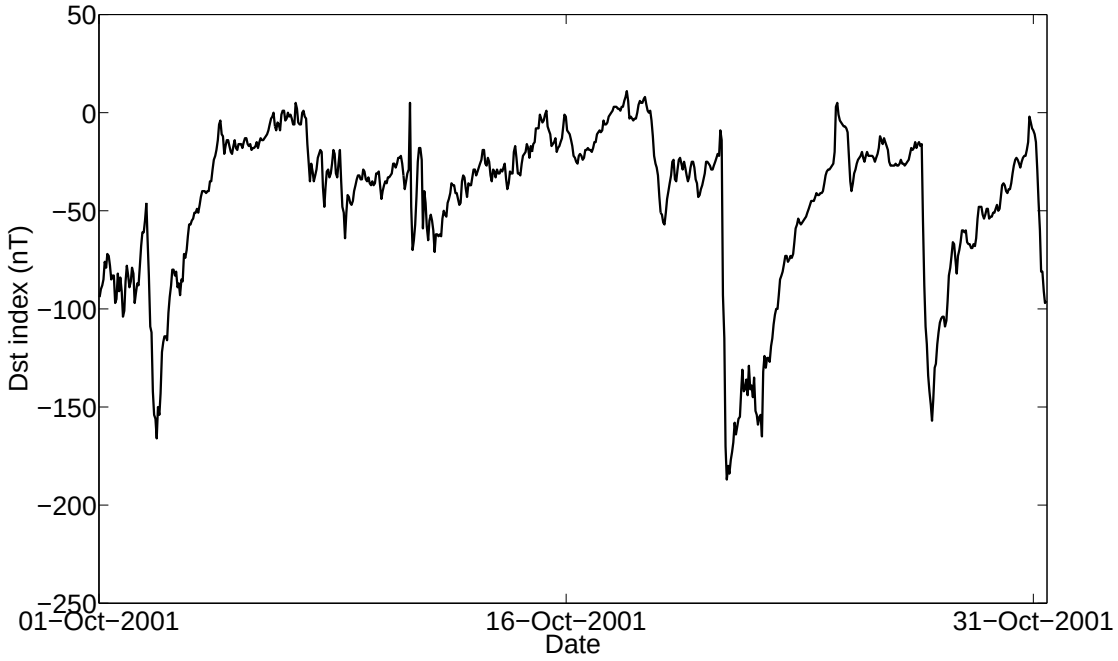

Figure 2.21: Hourly Dst index values for October 2001. Obtained from `http://wdc.kugi.kyoto-u.ac.jp/dstae/index.html`. Note the three major deviations, resulting from geomagnetic storms.

the VOs to amplify the magnetic effects resulting from small-scale temporal variations in the sampled data as the 'UT bias'.

The sectorial banding shown in in Figures 2.22 and 2.23 is for a month of highly active external fields. Figure 2.24 shows a month of Dst for a much lower field activity. As shown by Figure 2.25, sampling the low-activity Dst to the UTs of data inside the *Swarm* A VOs still produces high-order sectorial banding. This shows why the contamination present in the SHA residuals is such a persistent effect.

The tiling pattern shown in Figure 2.20 is much more spatially complex than the sectorial banding, exhibiting higher degrees and orders. This pattern is caused by VOs binning data in a manner which causes adjacent VOs to sample different amounts of the two available LT bands. This possibility was referred to in section 2.2.6 as 'LT-effects'. Here I will show how these effects can arise. Figure 2.26 shows the geographic distribution of data typical to *Swarm*

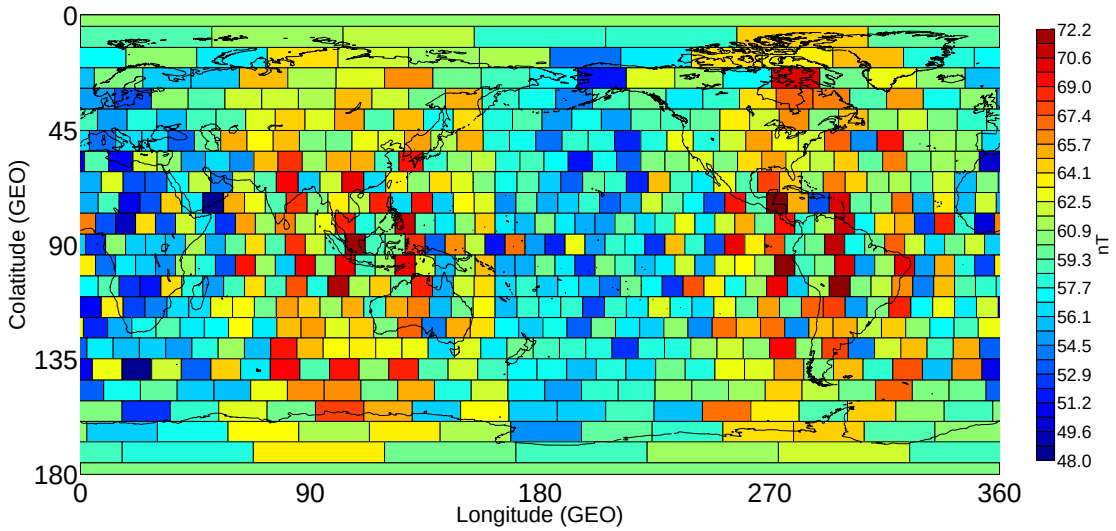

Figure 2.22: RMS Dst sampled values. Dst values of October 2001 (Figure 2.21) are sampled to the UTs inside each VO of *Swarm* A data for October 2001, then the RMS of each VO's Dst samples is taken.

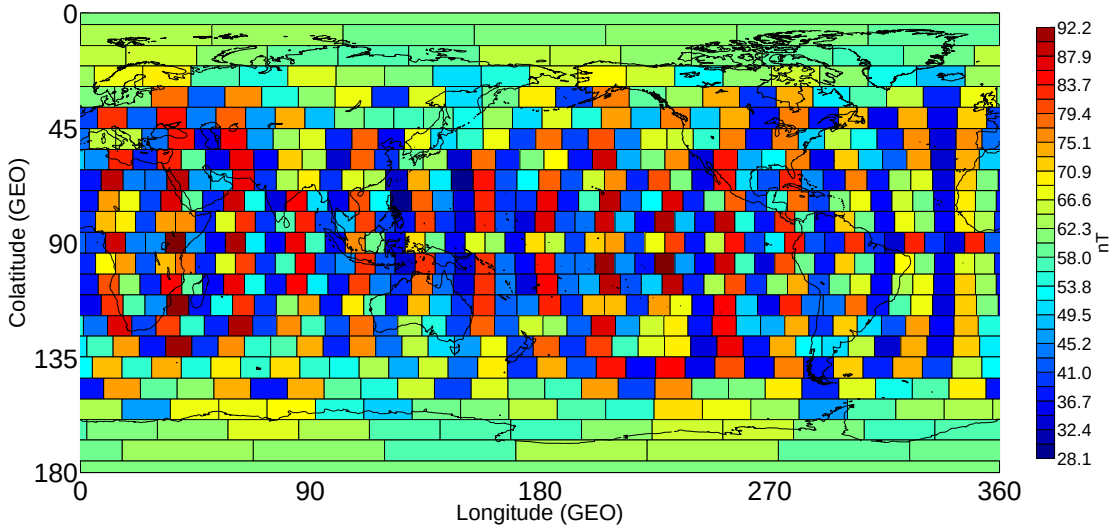

Figure 2.23: RMS Dst sampled values. Dst values of October 2001 (Figure 2.21) are sampled to the UTs inside each VO of *Swarm* C data for October 2001, then the RMS of each VO's Dst samples is taken.

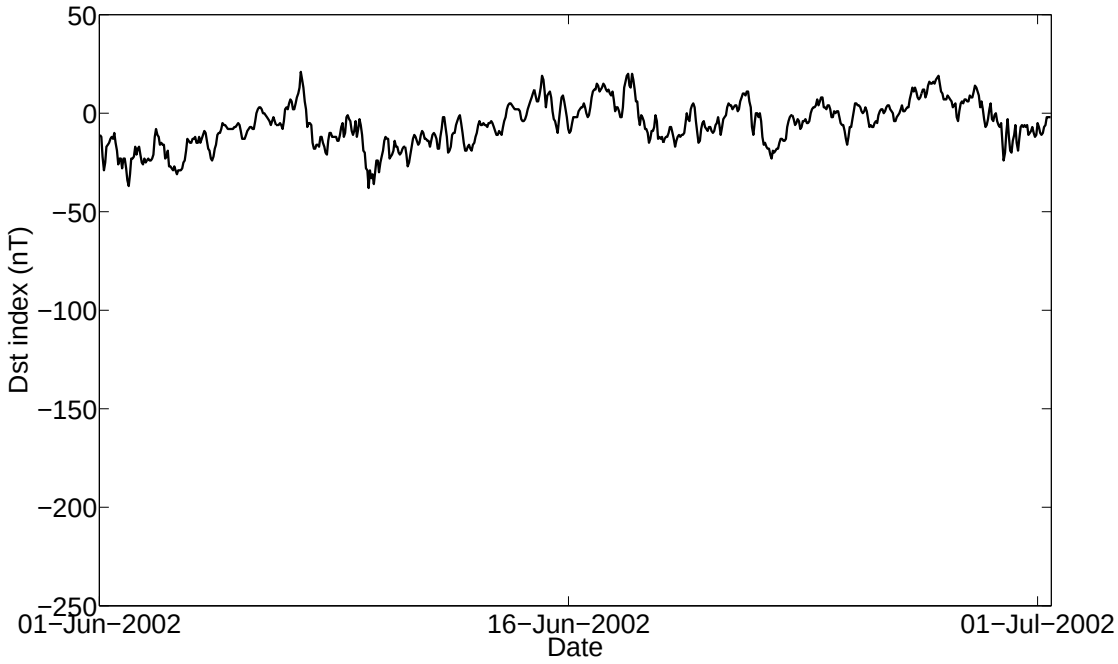

Figure 2.24: Hourly Dst index values for June 2002. Obtained from `http://wdc.kugi.kyoto-u.ac.jp/dstae/index.html`.

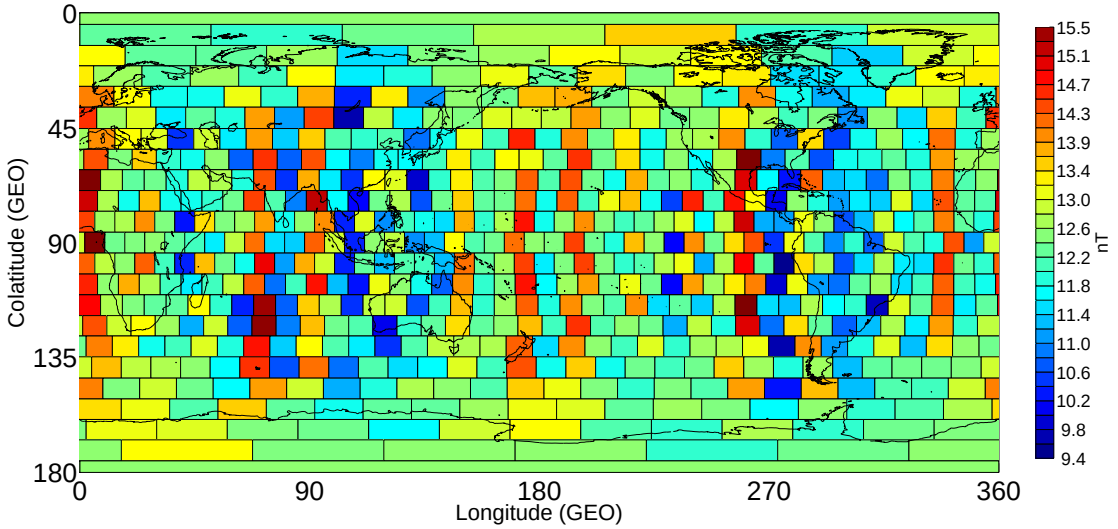

Figure 2.25: RMS Dst sampled values. Dst values of June 2002 (Figure 2.24) are sampled to the UTs inside each VO of *Swarm* A data for June 2002, then the RMS of each VO's Dst samples is taken. Note the difference in scale to Figure 2.22, though the sectorial banding appears similar.

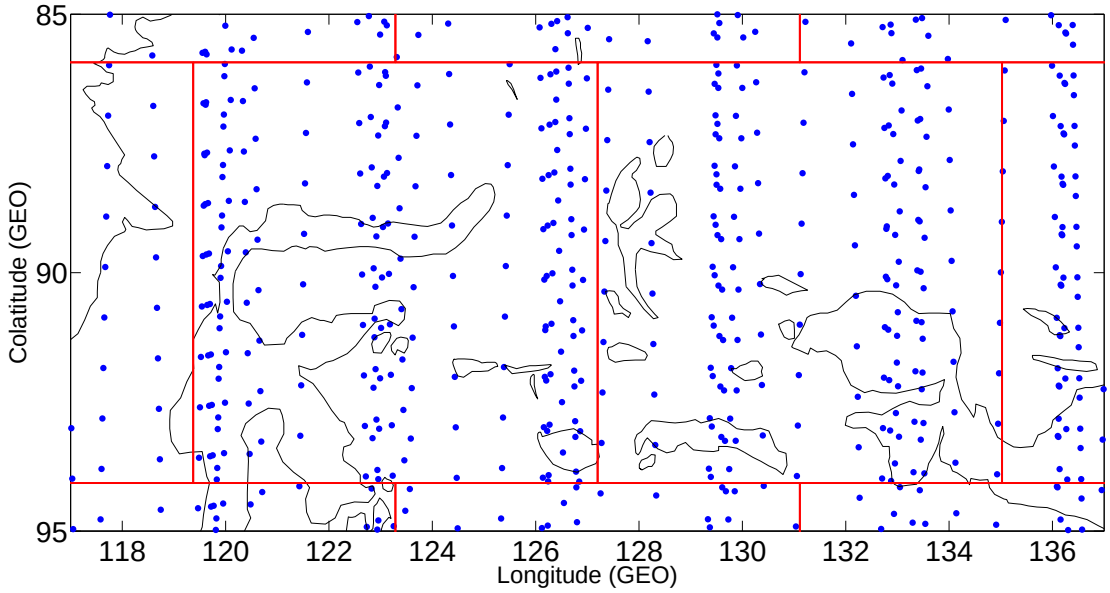

Figure 2.26: Geographic distribution of data in two equatorial VOs in 2001, October, for *Swarm* A. The distribution is not perfectly even, but is free from any systematic biases. The coastlines are of central and western Indonesia.

A equatorial VOs in October 2001. The distribution is imperfect, but free from systematic biases. Figure 2.27 shows the data distribution for the same VOs for *Swarm* A in September 2000 – these data are compressed longitudinally, as the satellite is coming out of an exact repeat orbit period. This is not necessarily damaging to the VO in itself, but each compressed 'band' of data points in Figure 2.27 is recorded at a different LT to the bands adjacent to it. Figure 2.28 shows the pattern of LTs recorded in two adjacent equatorial VOs in Figure 2.26 and Figure 2.29 shows the same for the data of Figure 2.27. Note that the amounts of each available LT band are sampled very differently between adjacent VOs in Figure 2.29 when compared to Figure 2.28. The difference between the number of data points from each band of LT in a certain VO can be used to reliably infer which of the two local times' magnetic signals will dominate the VO solution – this difference in point count is termed the 'LT bias'. The LT bias will only affect the global distribution of VO solutions if the magnetic environments in the two LT bands are significantly different (*e.g.* noon and midnight). As has been shown here, the resultant patterning is difficult to predict, and is exacerbated by situations of exact-repeat orbits.

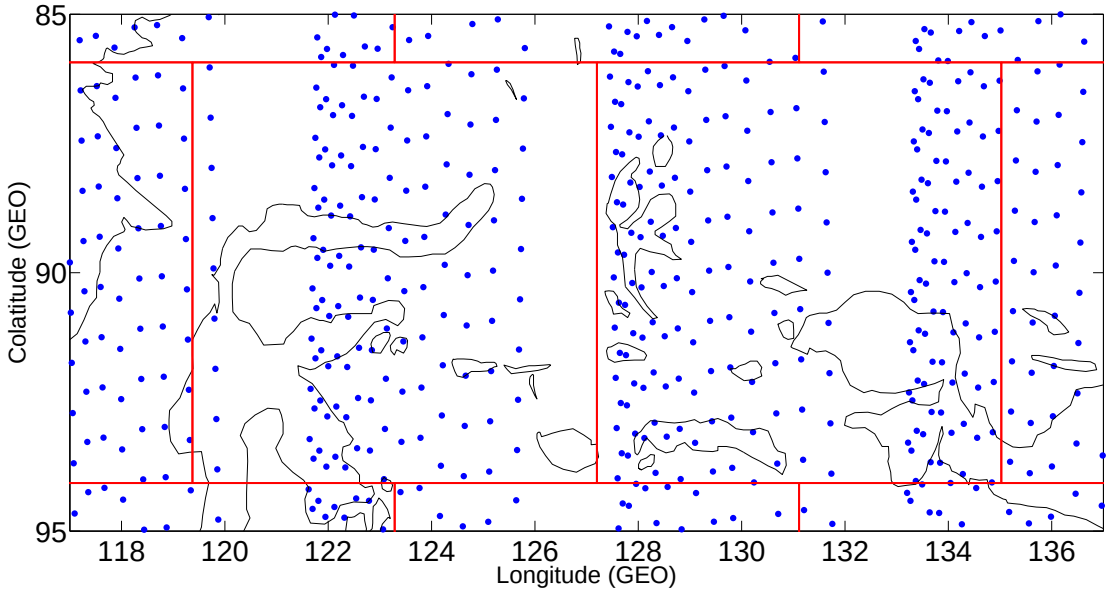

Figure 2.27: Geographic distribution of data in two equatorial VOs in 2000, September, for *Swarm* A. The distribution shows the systematic patterns resulting from a repeat-orbit situation. The coastlines are of central and western Indonesia.

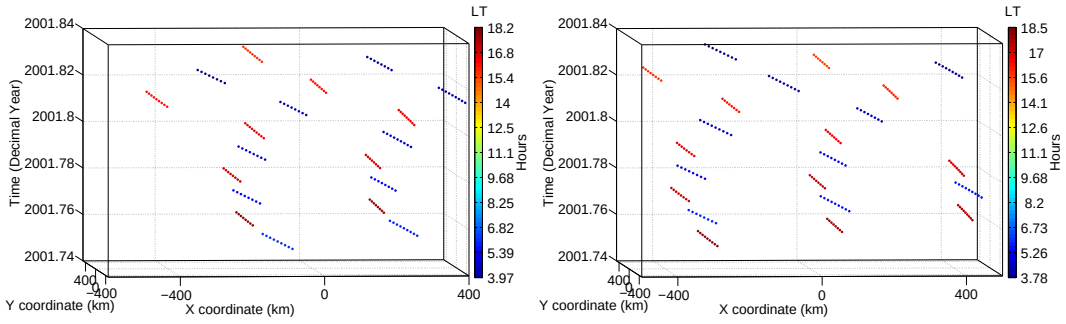

Figure 2.28: 3D distribution of data for two adjacent equatorial VOs in 2001, October, for *Swarm* A. The two horizontal axes (X and Y) are the target Cartesian coordinate system's horizontal axes. Since the altitude distribution of the data in the VO is not important here, the vertical axis is time (UT), intended to show the parts of the month being sampled by the VO. The colouration is LT, in order to show how the two bands of LT sampled by *Swarm* A for this month are represented. The figure is displayed with a slight azimuthal and elevation offset, to show the multiple data points making up each satellite 'pass' through the VO. Flight direction is roughly into and out of the page.

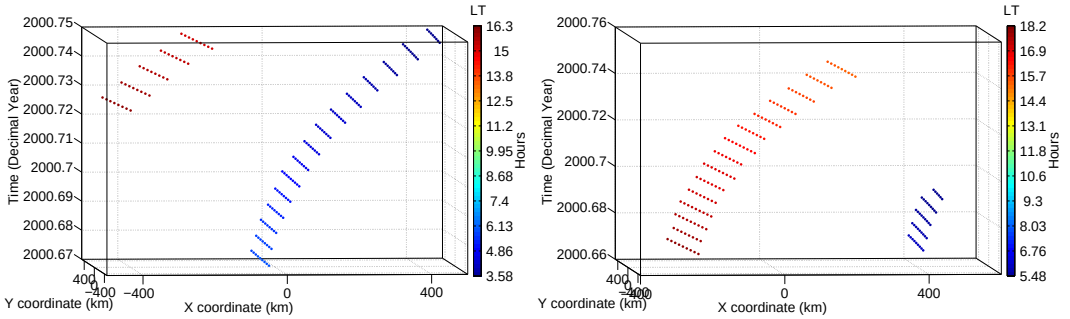

Figure 2.29: As Figure 2.28, but for September 2000.

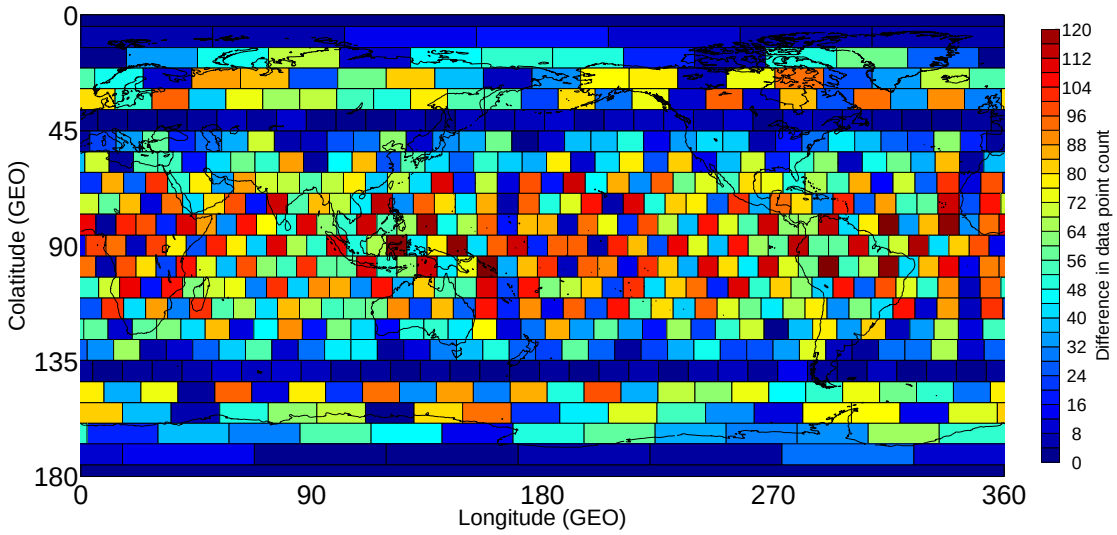

Figure 2.30: LT biases for each VO of *Swarm* A data for 2000, September. The LT bias is computed from the difference in data point count between the two LT bands recorded by each VO. It is an excellent analogue for the contamination patterns seen in Figure 2.20.

Figure 2.30 shows the global distribution of LT bias values for each VO in September 2000. The similarity of this pattern to the contamination distribution of the VO solutions in Figure 2.20 is very good, making the LT bias likely to be the cause of tiling-type contamination patterns. Conversely, the LT bias map of October 2001 for *Swarm* A shows no systematic trends, and hence the sectorial banding dominates this month of VO solutions.

The UT bias (sectorial banding) is more commonly present in the months of VO solutions than the LT bias (tiling pattern), but the latter is less easily predictable in its effects on the global VO solution distribution. The LT bias is part of the reason the utilisation of the constellation

geometry failed (further causes will be summarised in section 2.4). I have shown in this section that the VO sampling method is capable of aliasing fields which are both rapidly and slowly temporally variant. The study of the data distribution inside each VO also sheds light on why the IRLS method is not rejecting the external field value as outliers – each pass of the satellite through the VO records multiple data points, meaning that if a particular trace records a very magnetically active UT, the VO will have multiple instances of those high magnetic values and will be less likely to reject them as outliers. As stated by Chave et al. (1987), Huber weights "fall off slowly" for large residuals, and cannot prevent very high magnitude outliers from biasing the solution.

2.3.5 Isolating each source region's signal

The results presented so far have identified why the contamination is persistent in each month of E2Eplus data studied. It is likely that UT bias effects will impact real data in the same way as they affect the synthetic data. However, it is probable that the LT bias effects may be different (though still damaging) in real data, due to the simplification of the ionospheric and partial ring current fields in the E2Eplus model. This subsection assesses the impact of each modelled external field term on the VOs, and the SHA. I cannot assess this in real data as *Swarm* has not launched at the time of writing, but will attempt to suggest instances where the simplicity of the modelled external fields is damaging in assessing how much of it is aliased into the global SHA model.

Figure 2.31 shows the global averages of the VO solutions for each month of E2Eplus data, for varying input types. The black line is for all external field terms (main field and SV terms are still not being included in the analysis). The red line is magnetospheric and induced terms, blue line is toroidal terms and the green line is ionospheric and induced terms. The magnetospheric terms represent the largest magnitude external field signal in the VO solutions. Toroidal terms are second in terms of signal strength, though I only include these for completeness as the modelled toroidal fields are likely inaccurate in the E2Eplus model. The smallest contribution is from the ionospheric terms, which have a clear 4.5-monthly oscillation related to the

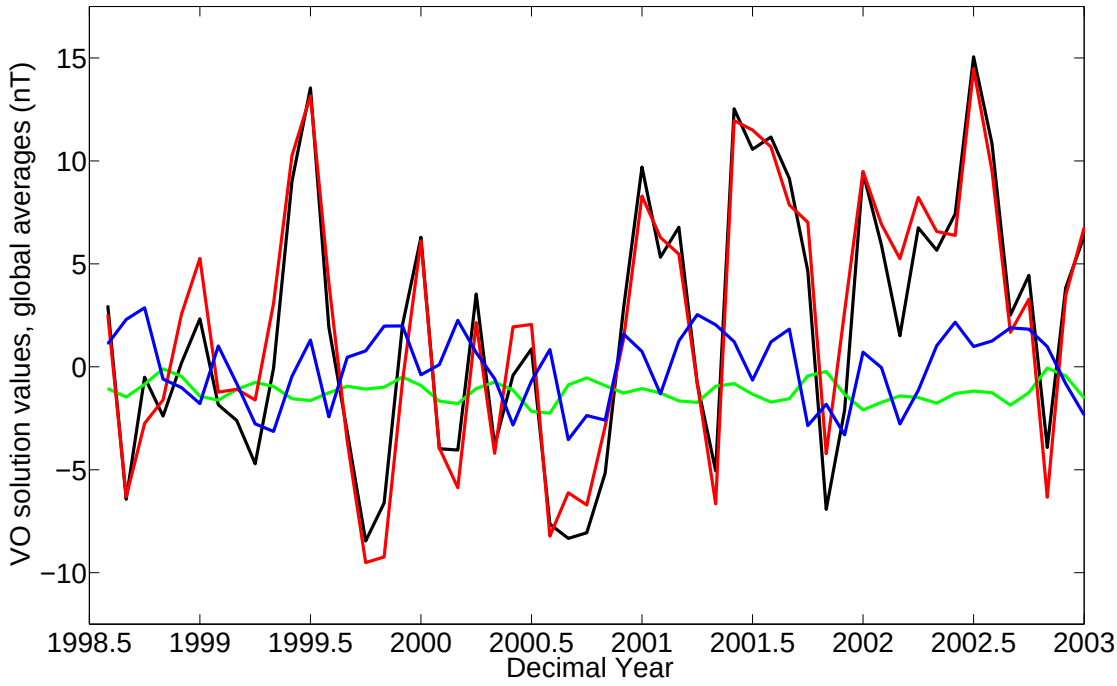

Figure 2.31: Global averages of VO solutions of *Swarm* A data, for input data with varying source-region contributions, for all 54 months of E2Eplus data (UT frame). Black line is all field contributions (less SV and main field). Red line is magnetospheric and induced terms. Blue line is toroidal terms. Green line is ionospheric and induced terms.

time taken by the *Swarm* A satellite to sample all LTs.

Figure 2.32 shows the VO solutions for *Swarm* A with only ionospheric and induced terms as a smooth surface for October 2001. The distribution of VO solutions shows some sectorial banding near the equator, but is globally very smooth – this is the 'smeared' signal from a month's worth of samples of the same two narrow LT bands. Comparing Figures 2.32 and 2.19 it is seen that (although this signal is for different satellites), the ionospheric signal appears to have a strong effect on the SHA model's data prediction. In the real data a similar effect could be expected, though one which will be lessened in magnitude as the real ionosphere signal would not be so smooth when sampled in this manner. Figure 2.33 shows the VO solutions for *Swarm* A with only magnetospheric and induced terms as a map for October 2001. The major part of the sectorial banding is due to the magnetospheric signal, as expected from the Dst-sampling experiments discussed above.

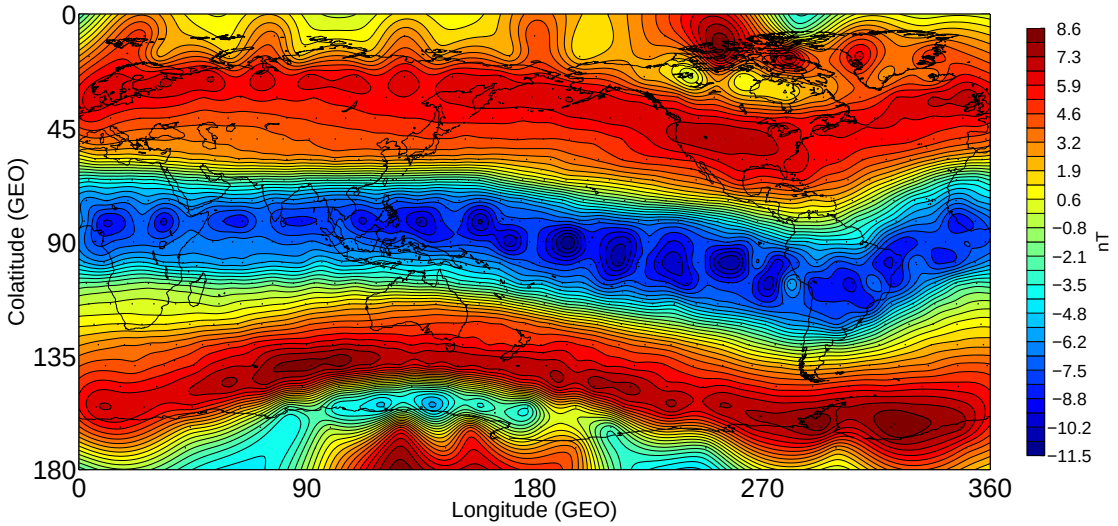

Figure 2.32: VO solutions for the θ-component of the magnetic field using *Swarm* A data, October 2001. Only ionospheric and induced terms are included.

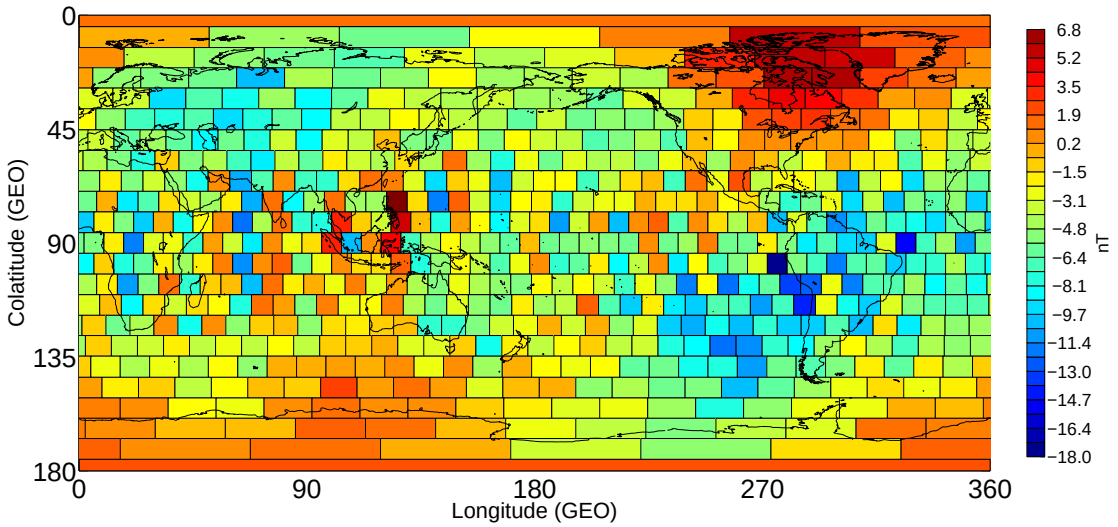

Figure 2.33: VO solutions for the θ-component of the magnetic field using *Swarm* A data, October 2001. Only magnetospheric and induced terms are included.

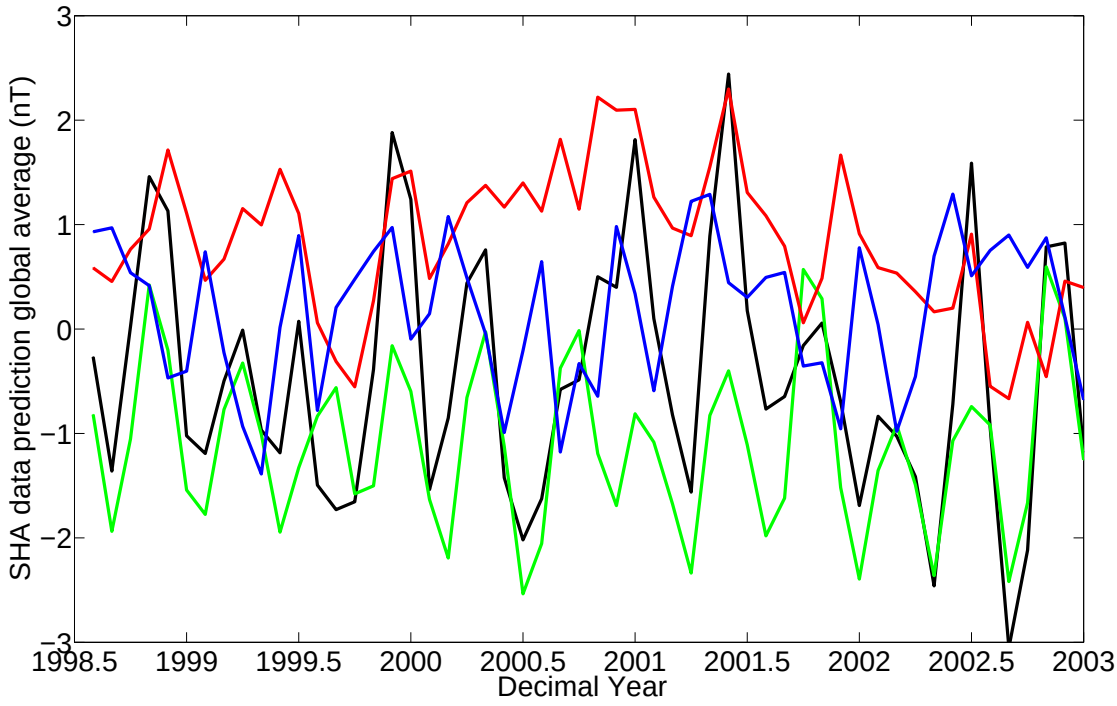

Figure 2.34: Global averages of residuals of *Swarm* A VOs to monthly global degree 14 SHA models, for input data with varying source-region contributions, for all 54 months of E2Eplus data (UT frame). Black line is all field contributions (less SV and main field). Red line is magnetospheric and induced terms. Blue line is toroidal terms. Green line is ionospheric and induced terms.

The global mean SHA data predictions (to the VO locations) are shown in Figure 2.34, for a range of input data types. The line colouration denotes the same sources as in Figure 2.31. The tendency of the magnetospheric fields to be aliased more strongly by the VO sampling method increases their spatial complexity in the VO solutions, thereby reducing their contribution to the SHA model relative to the other external field terms. This indicates why the (real and synthetic data) SHA residuals have a strong correlation with the Dst index, which represents the dominant magnetospheric field source region. The black line is the global mean SHA residual from all field sources (without the main field and SV). I have shown in Figure 2.11 that (albeit in the SV frame) its variation is damaging to studies of the SV. It appears from the magnitudes of the time series in Figure 2.31 that enough of the external field signal from each of the three sources shown is aliased into the SHA model to present a significant bias in estimating the SV, though this has been demonstrated only for the E2Eplus model.

2.3.6 Attempts to mitigate the effects of the contamination

Although the initial focus of this study on synthetic data was expected to progress to real data in due course, at this stage in the study it became apparent that the *Swarm* mission would not launch during the project. The lack of real data places a limitation on what can be said about the effectiveness of the proposed contamination mitigation strategies I discuss in this section, but I discuss effects which should be pertinent to both synthetic and real data.

Attempted fixes

I have shown in section 2.3.2 that utilising the *Swarm* constellation geometry has no significant effect on reducing the contamination, in either the VOs or the global SHA model. In this section I discuss other attempts at mitigating the contamination in the VOs, none of which were fully successful. The discussion here is not specific to constellation satellite missions, but is aimed at addressing some perceived shortcomings of the VOs which used the synthetic data from all *Swarm* satellites. I have noted earlier that the rejection of the input data values by the IRLS method was not reducing the contamination in the VO solutions. In section 2.3.4 I considered that this could be due to the satellites recording multiple data points in each VO per fly-through, which in times of high field activity might impact the 'identification' as outliers of even strongly biased (by external fields) magnetic data. I therefore tested a method of selecting the input data for a VO randomly from the data it has 'sampled', and computing the local solutions from this random subset. However, this did not significantly reduce the contamination. I refined the method, attempting to tailor the statistical distribution of the input data to be the best fit to a Gaussian distribution (via iterations through randomly-selected subsets), but again, the method was not successful.

A partially-successful change to the VOs was that of increasing the VO data bin's temporal length. Note that the *Swarm* constellation samples at maximum four distinct bands of LT per month. However, a single satellite (when both ascending and descending orbital nodes are considered) can sample all LTs in roughly 4.5 months. I thus tested expanding the temporal length of the VOs such that the sampled data covered all local times. The effect of this step on

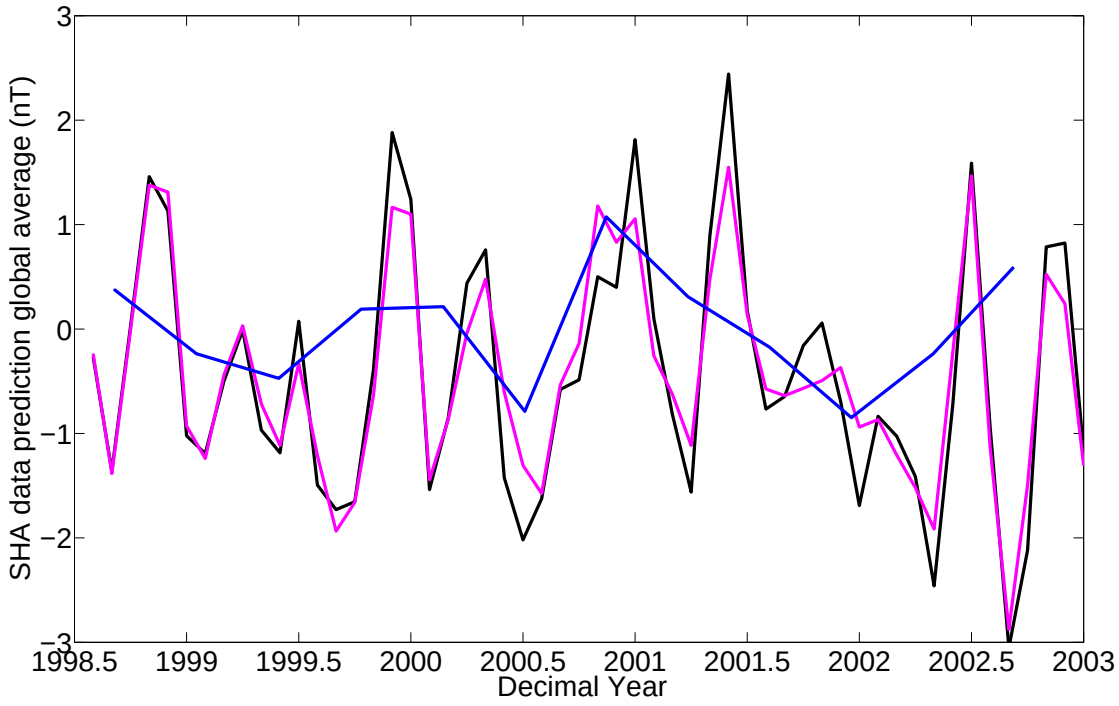

Figure 2.35: Global means of θ-component data predictions at 648 VO locations of degree-14 SHAs of various VO methods applied to *Swarm* synthetic data. Black line is 1-month VO method with *Swarm* A data. Magenta line is the 1-month VO method with all *Swarm* satellite data. The blue line is the 4.5 month long VO method, with *Swarm* A data. All time series have had a linear trend removed, in order to exhibit small temporal-scale variations.

the global mean data predictions (UT frame) is shown in Figure 2.35. The black line is 1-month VO method with *Swarm* A data, the magenta line is the 1-month VO method with all *Swarm* satellite data, and the blue line is the 4.5 month long VO method, with *Swarm* A data. All time series have had a linear trend removed, in order to exhibit small temporal-scale variations. Note that the reduction of the contamination by the 4.5 month-long VO method is minimal beyond the smoothing caused by using longer VOs. This indicates that improving the representation of the LT fields alone is not sufficient to reduce the contamination significantly, and that the result of Beggan (2009) that the contamination is largely due to the symmetric ring current, is correct. Were the real *Swarm* data available I would have attempted a more sophisticated set of tests involving the temporally-extended VOs, but this is beyond the scope of the study. In the following section, I discuss the steps required to produce less-contaminated VO solutions.

Suggested fixes

For the VO method to work effectively, it must remove the contamination from the LT-arranged fields and the symmetric ring current, avoid the LT bias effects and ensure that the seasonal signal is removed in the SV frame. Here I suggest the steps required to achieve these constraints. Real satellite data are required from at least one satellite in polar LEO (with the others orbiting in a different LT sector). The use of multiple satellites avoids issues resulting from repeat-orbit effects and ensures a good coverage of the temporal variations at different LTs. The data must be corrected for the magnetospheric contributions. The magnetospheric correction can be approached in one of three ways: firstly, a model of the symmetric ring current can be removed from the input data. Secondly, the satellites providing the VO input data could be used to solve for the magnetospheric dipole (mean signal from half an orbit in geomagnetic dipole coordinates), producing two 'magnetospheric plus mean magnetic environment' corrections per orbiting satellite (a simpler version of the approach discussed in Sabaka et al. (2002, page 45)). The magnetospheric dipole solution is currently being developed for *Swarm* by Brian Hamilton at the British Geological Survey. Thirdly, a meridional chain of permanent observatories could be used (if properly weighted for latitude) to solve for the magnetospheric dipole in the same way, producing an LT-specific external field correction every hour of LT. The corrected data should be binned in a series of 648 tesseral VOs of temporal length 4.5 months (or as long as is taken for all satellites to sample all LTs). Prior to computing the local solutions, the data should be winnowed such that each LT sector (with sector width dependent on the number of data points available) has an even contribution to the VO solution, to avoid LT-bias effects. The use of multiple satellite aids in the additional data coverage required for the winnowing.

The SV can then be computed from the VO solutions. The use of 4.5-monthly VOs means that the SV cannot be computed directly at 12 monthly intervals without temporal interpolation, which is an overcomplication. However, the seasonal signal must still cancel out in the SV frame. Note that 8 'cycles' of a 4.5-monthly period is equivalent to 36 months, meaning that the SV can be computed at a three-year period and in doing so, the seasonal signal effects are removed as well as a consistent LT-arranged fields contribution. The SV would be smoother

than that obtained from the 12-monthly first time difference and would have a temporal resolution of 4.5 months rather than 1 month, but should have greatly reduced external field effects. This approach is not suitable for application with the synthetic data, as the external field model is likely to be too simplistic for the outcome to tell us anything useful.

2.4 Discussion

Beggan (2009) showed that the VO contamination was principally due to the aliasing of the symmetric ring current (apparent also in Figure 2.31) but that correcting for this signal with the CM4 model introduced an additional signal into the VOs. It therefore seemed likely that the poor sampling of the LT-arranged fields by a single satellite was contributing significantly to the contamination in the SHA residuals. Here, I have confirmed that supposition. However, I have shown that the causes of the aliasing were more wide-ranging than initially expected, and that using the VO method to sample the external fields with more satellites does not trivially aid the separation of internal and external fields.

From the results of sampling the Dst index to the times of the data used in the VOs, I have shown that the most common aspect of the contamination (the sectorial banding, resulting from aliasing of rapidly variant external fields) can be replicated with a globally uniform signal. The replication of this signal persists even in magnetically quiet months. The contamination is much more persistent than initially expected, and is likely to bias any SV estimate unless properly mitigated. From the results in the previous section, it is likely that the signal in Beggan (2009)'s study introduced by CM4 was the LT bias, *i.e.* the aliasing of the slowly-varying LT-arranged magnetic fields (the ionospheric and low-activity partial ring current contributions). It was not initially expected that the VOs would alias these signals. I have shown in section 2.3.2 that utilising the *Swarm* constellation geometry has no significant effect on reducing the contamination, in either the VOs or the global SHA model. The reasons for the lack of reduction are now apparent. The initial assumption that the LT-arranged fields were contributing significantly to the contamination was correct, but this was not due to them being inadequately sampled in the LT frame, but rather being due to them being aliased via the LT bias effects.

Adding an additional LT-sector of sampling locations had no effect in reducing the LT bias, and hence had no effect in reducing the parts of the contamination which were due to LT. The aliased representations of the LT-arranged fields were not separated out into the residuals by the global SHA, and they do not cancel out when the first difference is taken (SV frame), leading to a minimal benefit from applying the *Swarm* constellation geometry to the VOs.

An additional issue with the use of the constellation in the VOs (which I have not previously discussed) is that the *Swarm* geometry is intended to provide simultaneous sampling of four sectors of LT at once. The *Swarm* mission certainly provides this, but the VOs cannot make use of the simultaneous sampling. Consider the setup of a single VO at 3.5 years into the *Swarm* mission, using data from all three satellites for one month. At a given initial time, the geographically fixed VO will record a trace from *Swarm* A (and *Swarm* B) at a certain LT. *Swarm* C is, at this time, observing the magnetic field at a different LT sector. In order for the VO to sample that LT sector, we have to wait for the Earth to rotate underneath the *Swarm* sampling shells until the VO can sample the data that *Swarm* C is recording. This takes 6 hours UT, by which time the external fields (whether LT-arranged, or zonally continuous in LT) will have varied, meaning that a different signal is sampled. Therefore, another major reason that the constellation geometry did not reduce the contamination is that it does not provide enough additional sampling locations in LT to mitigate the UT bias, which is related to the aliasing of rapidly-varying external field (such as the symmetric ring current). My study indicates that (in the E2Eplus model), the Nyquist period of the external field variations (when sampled at a single geographically-fixed location) is less than 6 hours UT.

The VO method removes the benefits of the long-period recording times which LEO satellites provide. Since here the (effectively LT-fixed, during one month) satellites are used to study small-scale, geographically-fixed phenomena, the maximum sampling rate allowed by the constellation geometry for a single GEO-fixed location must be considered. The maximum sampling rate is limited by the rotation rate of the Earth (assuming polar-orbiting satellites). However, the satellite orbit velocity, the longitudinal width of the phenomenon being sampled

(here 900 km for the VO), and number of satellites also have an effect on what can be sampled. In general, to properly represent a magnetic field signal as a time series at a GEO-fixed location by a polar-orbiting satellite constellation, the separation of the satellites in LT must be less than the Nyquist frequency of (the bulk of) the magnetic field signal in UT. Again, this assumes that the satellite's precession rate in LT is negligible on the scale of 24 hours – equatorially-orbiting satellites will have an advantage in rapidly sampling all local times.

It is expected that the Nyquist frequency (*e.g.* when sampled at a geographically-fixed location) of the external fields will increase with activity level. This is the reason I suggested in section 2.3.6 that the VO input data be corrected for the variations of the magnetospheric fields, under the assumption that the ionospheric fields are less rapidly variant. An accurate forward model for the magnetospheric field variations at high activity levels on the Earth's surface does not currently exist, but our understanding of the Earth's external fields is continually improving, aided by the launch of missions such as the Radiation Belt Storm Probes (RBSP)[5]. In the resource exploration industry, the use of temporally discontinuous track-lines of magnetic data is common. An alternative to requiring extra LEO Earth satellites in studying GEO-fixed phenomena could be to use a sufficiently sophisticated magnetospheric model to provide 'cross-track' levelling of the satellite passes (*e.g.* recorded by each VO). Thus if the magnetic fields cannot be sampled densely enough (temporally), then the recorded traces can be stripped of the magnetospheric signal in near-real-time.

2.5 Summary

I initially assumed that identifying the causes of the contamination affecting the VOs would suggest contamination-mitigation strategies that would be directly applicable to all parameterisations of the external field, that is, that any errors identified in the VOs would affect any study in which external field effects are important. It seems that the VO contamination is in fact a mixture of effects which impact all studies, and those which only affect the VOs. The

[5]RBSP was launched on 30th August 2012. Mission statement at `http://www.nasa.gov/mission_pages/rbsp/main/index.html`.

LT bias will only affect studies which seek to resolve GEO-fixed phenomena via binning of satellite data. However, the UT bias (the undersampling of rapidly-varying external fields) is a more wide-ranging effect, with the potential to affect a broad range of magnetic field studies. I was initially concerned that the *Swarm* constellation would not be able to resolve the external fields as well as other studies have suggested. From the discussion in sections 2.3.4 and 2.4, it is apparent that the problem lies in how the VO method samples the constellation data, rather than how the constellation samples the external fields. Whilst the VO contamination cannot be simply removed, based on the discussion in this chapter I have suggested strategies for its mitigation, though these synthetic data strategies might require modification when applied to real data, owing to their greater complexity.

This study has highlighted two main areas for future research. Firstly, I have assumed in my use of the *Swarm* constellation geometry, and in my suggestions (section 2.3.6) for modifications to the VO method, that the external field variations in local time can be fully sampled by four (slowly precessing) sectors of local time. However, this assumption may not hold true in all situations. After the delay in the launch of *Swarm*, the temporal and spatial variations of the external fields were identified as an area of the magnetic field requiring extra research, and the EOF study (chapter 4) was commenced in order to resolve the dominant spatiotemporal patterns of the external fields. Secondly, the assumption of the VO method that the LEO altitude electric currents were invariant with altitude is incorrect. The Ampère's integral study (chapter 3) was undertaken in order to resolve, for the first time, electric currents at LEO satellite altitude without the need to parameterise or average-out magnetic contributions from regions above or below the satellite's altitude.

Chapter 3

Ionospheric mid-latitude electric current density inferred from multiple magnetic satellites using Ampère's integral

3.1 Glossary

Table 3.1: Glossary of terms used in this chapter

Abbreviation	Explanation
Line element	Part of an integral loop – the distance between two successive data points
Contribution value	A single line element's integral-contribution value which, when summed, will give the current flowing through the integral area
Altitude imbalance	The difference in altitude in the two radial line elements connecting the two arcs in each integral loop
VO	Virtual Observatory (plural VOs: Virtual Observatories), described in chapter 2
LT	Local time
UT	Universal time
GEO	Geographic coordinate system (Hapgood, 1992)
QD	Quasi-Dipole coordinates (Richmond, 1995)

Acknowledgements

The work in the Ampère's integral chapter was done by me. I presented the preliminary results of my analysis at two international conferences, during which discussion with Hermann Lühr, in particular, added valuable expertise to the approach I used. Following these conferences I spent a month at DTU Space in Copenhagen, Denmark, working with Nils Olsen (mainly on a separate study). During this time Nils provided me with access to the synthetic *Swarm* constellation data generation program, and related code to change the modelled attitude of the synthetic satellites. The outcome of this addition to the Ampère's integral study increased my trust in the electric current density estimates. Upon returning to the UK I worked with Anasuya Aruliah and Timothy Spain at UCL, who provided data predictions of current density from the CTIP model, allowing me to make broad comparisons with that first-principles ionospheric model, again improving trust in the results of my study. Thanks also go to Roderick Heelis, for an illuminating discussion on equatorial ionospheric electric current structure. Above all, thanks go to my supervisors, who were immensely supportive throughout the study, not to mention very patient in reading through all of the numerous versions of the paper Shore et al. (2013) based on this work. My supervisors also helped with proofreading this chapter.

3.2 Motivation

Magnetic field research has greatly benefitted over the past decade from an abundance of high-accuracy satellite-sampled data obtained at low-Earth orbit (LEO). The spacecraft fly through the magnetised plasma of the ionosphere in which a complex array of *in situ* electric currents flow in both the day and night sectors. The currents contribute to the measured magnetic field, violating the assumption of measurement in a source-free region – a requirement of magnetic field representations which adopt a scalar potential. In the Virtual Observatory (VO) study (chapter 2), I queried the validity of the assumption of the method of Mandea & Olsen (2006) that the electric currents at LEO altitude do not affect the VO solutions, but this approach is common. In standard spherical harmonic analysis (SHA) methods applied to satellite magnetic data, data selection techniques (*e.g.* retaining only nightside data) are commonly employed to

lessen the effect of the unwanted magnetic contributions from *in situ* electric currents, after which the influence of the currents is assumed negligible. Prior to the Decade of Geopotential Field Research (Friis-Christensen et al., 2009), this assumption was appropriate for magnetic field models based largely on observations made on or near the surface of the Earth. However, to make the fullest use of the abundance of high-precision magnetic data collected by CHAMP, Ørsted and SAC-C, the LEO distribution of electric currents (and the associated magnetic fields) must be better accounted for, requiring an improved understanding of their distribution and magnitudes.

Much work has been published on the description and quantification of the LEO altitude currents with single-satellite measurements, for example by Olsen (1997), Lühr et al. (2002), Maus & Lühr (2006) and Lühr & Maus (2006). However, the currents at LEO altitude are weaker than those in the lower ionospheric or higher magnetospheric regions (conveniently, given that their contributions are typically unwanted) . When using a single satellite to estimate LEO altitude electric current density, it is difficult to avoid bias from inclusion of contributions from regions outside the satellites' altitude range. This can be ameliorated in part by making geometrical assumptions about the regions being avoided, as in, for example, Juusola et al. (2007), but the error associated with such methods may be comparable in magnitude to the signal of the currents at LEO altitudes.

In this chapter I describe a method of using near-simultaneous overflights of the Ørsted and CHAMP satellites to define a closed circuit for an application of Ampère's integral law to magnetic data, in order to estimate total current flow in the region between the two orbital paths on an event-by-event basis (*i.e.* separately for each discrete overflight). I resolve, for the first time, multi-satellite-estimated trends in the zonal mid-latitude ionospheric electric current density at a range of local times (LT) and phases of the solar cycle, obtained without the need to parameterise or average-out magnetic contributions from regions above or below the satellites' altitude. The orbits of the Ørsted and CHAMP satellites are near-polar-crossing with altitudes of 650–860 km and 350–460 km respectively (Neubert et al., 2001; Reigber et al.,

2002). Six years of mutually available vector magnetic data allows overlaps spanning the full 24-hour range of LTs twice.

My major research questions in this study are as follows:

- Can the electric currents at LEO altitude be adequately resolved by processing existing satellite magnetic data in a constellation configuration which was not designed for this purpose?
- Could the same approach work with the *Swarm* constellation geometry?
- Do the results have implications for the typical approach of selecting magnetic data from the nightside and considering it free of *in situ* currents?
- Are the resolved current density patterns also seen in physics-based models of the ionosphere?

As I discuss in more detail in section 3.5, the use of Ampère's law integral on isolated overflights of Ørsted and CHAMP precludes a global perspective which is sustained through time, so I assess in detail what could cause the observed trends in the current density estimates. Much of the next section will be devoted to an overview of the pertinent ionospheric electrodynamics, how they lead to accumulations of plasma, and where stronger currents can be expected to be observed as a result. Although it is not possible to state uniquely which current system is dominant in my observed trends (shown in section 3.5), the following review of the controlling processes in the area of interest will allow educated guesses to be made.

3.3 Ionospheric electrodynamics at low latitudes

The ionosphere is the conducting region in the Earth's atmosphere between 60 and 1,000 km (Campbell, 2003, page 74) with the protonosphere (plasmasphere) above. The upper boundary ('topside') of the ionosphere is rather diffuse, as well as being laterally variable in altitude, but the typical definition of 1,000 km as its limit will suffice, as this is above the range of Ørsted orbital altitudes. The ionosphere spans both the mesosphere and thermosphere, and is stratified

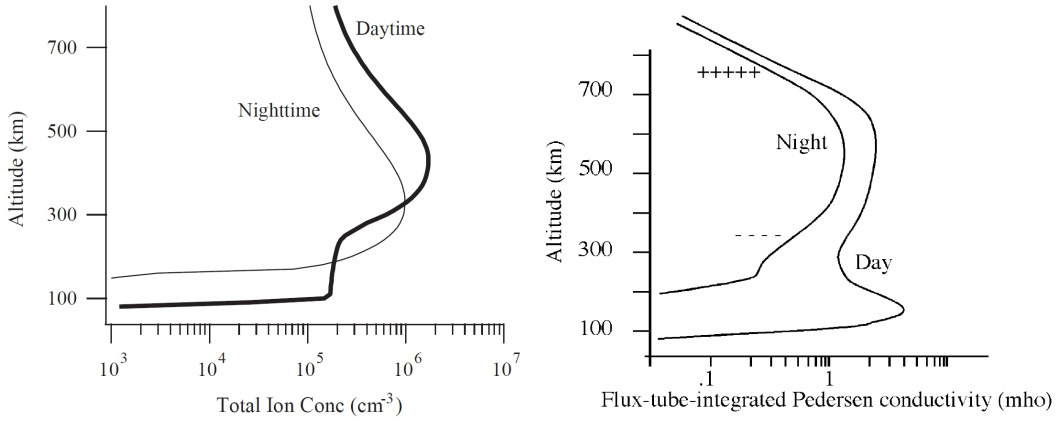

Figure 3.1: Generalised altitude distributions of ionospheric ion density (a) and resulting Pedersen conductivity distribution (b), from Heelis (2004). Note the difference between the two distributions, particularly in the dayside E-region at 100 km altitude. The $+$ and $-$ symbols in the right-hand image indicate the charge-build-up associated with the F-region Cowling effect, described in more detail in section 3.3.2.

in terms of plasma density into the E- and F-regions, which occupy altitudes of 90–140 km and 140–1000 km respectively (Kelley, 2009). I focus on the F-region in this study, which fully encompasses the orbital altitude ranges of both Ørsted and CHAMP.

The altitude profiles of the mean mid-latitude ion density (hence, plasma density) and Pedersen conductivity are shown in Figure 3.1. Ionospheric conductivity is highly anisotropic, and strong currents in one direction can have minimal effects in other directions Forbes (1981). As described by *e.g.* Rishbeth (1988), the Pedersen conductivity relates the electric field, in the direction perpendicular to the magnetic field, to the current flow in that direction. The Hall conductivity describes current flow in the direction perpendicular to both electric and magnetic fields. As stated by Heelis (2004), Hall conductivity is negligible in the F-region. The currents I resolve in this study stem from the Pedersen conductivity distribution. Utilising the overflight configuration of the Ørsted and CHAMP satellites allows resolution of predominantly zonal (east-west) electric currents flowing in the region between them at mid-to-low latitudes. The geometry of the overflights precludes analysis of the currents at high latitudes, as I discuss further in section 3.4. At F-region altitudes, the major zonal electric current sources are due to the action of Lorentz, gravity and pressure gradient forces, with the full distribution of current

density given by (Lühr et al., 2008b; Alken et al., 2011)

$$\begin{aligned} \mathbf{J} &= \underline{\sigma}\,(\mathbf{E}+\mathbf{U}\times\mathbf{B}) \\ &\quad +\{N_e m_i \mathbf{g}\times\mathbf{B} - k\nabla\left[(T_i+T_e)N_e\right]\times\mathbf{B}\}\,\frac{1}{B^2} \end{aligned} \tag{3.1}$$

where $\mathbf{J}$ is the current density, $\underline{\sigma}$ the conductivity tensor (Forbes, 1981), $\mathbf{E}$ the electric field, $\mathbf{U}$ the neutral wind velocity, N_e the electron density, m_i the ion mass, $\mathbf{g}$ the gravitational acceleration, k Boltzmann's constant, T_i and T_e the ion and electron temperatures, and $\mathbf{B}$ the ambient magnetic field with magnitude B.

This study uses the geographic (GEO) system of spherical polar coordinates (r,θ,ϕ), where r is the length of the radial position vector from Earth's centre (and 'altitude' is the part of r above the reference sphere of radius 6371.2 km), θ (colatitude) is the angle between the radial vector and the Earth's rotation axis, and ϕ (longitude) is the angle in the equatorial plane between the radial vector and the Greenwich meridian (*e.g.*, Hapgood, 1992).

The first term in equation (3.1) is the combination of the electric and magnetic Lorentz forces, modulated by Ohm's law via the conductivity tensor (Rishbeth, 1988; Kivelson & Russell, 1995), and describes the major ionospheric dynamo system currents such as Solar-quiet (Sq) and the equatorial electrojet (EEJ) (Alken et al., 2011). These current systems result from the tendency of the ionosphere to build up electric charges at the gradients between regions of contrasting conductivity, thus creating secondary 'polarisation' electric fields in order to keep the global divergence of the current (inside the ionosphere) zero, as required by Poisson's law (Rishbeth, 1997). The polarisation electric fields in turn cause secondary (feedback) currents (Amm et al., 2011). Cowling (1932) was the first to recognise this property of the ionosphere. The two parts of the second term in equation (3.1) describe the gravity and pressure gradient electric currents. Rishbeth (1988) described the electric forcing, while Goldberg (1965) discussed the gravity and pressure driven currents (neglecting the effect of neutral air winds). The

gravity and pressure gradient currents only make a significant contribution to the total current at magnetically quiet times.

Whilst the gravity and pressure gradient current directions are relatively straightforward to predict, the direction of the (typically stronger) Lorentz current is not simply determined, and is dependent on a number of competing, time-variant factors, even before divergence-equalising feedback currents are considered. As stated above, the geometry of the satellite overflights constrains my analysis of the electric currents mainly to those flowing in the zonal direction. A global determination of the expected trends in the solely zonal-direction electric currents would be very useful, but it is beyond the scope of this project. In section 3.5.2 I use an ionospheric model to provide an assessment of the relative contribution of each term in equation (3.1) to the zonal current density. Here, I focus on the expected trends of current density intensifications, via a description of the electrodynamics controlling the distribution (in dip latitude and LT) of plasma density, which is a key factor in modulating each term in equation (3.1).

In the mid- and low-latitude ionosphere, electric, gravity, plasma-pressure and neutral-pressure forces push the plasma across magnetic field lines. Whilst the ionosphere as a whole can reasonably be defined as horizontally stratified, the electrodynamic coupling of the E- and F-regions produces complex structure and marked meridional gradients in the plasma distribution. In the area of interest, the largest of these structures is the Appleton anomaly (Appleton, 1946), also known as the Equatorial Ionisation Anomaly (EIA), characterised by two 'crests' of enhanced plasma density flanking the magnetic dip equator at tropical latitudes (Alken & Maus, 2010; Alken et al., 2011). The EIA is the main ionospheric phenomenon resolved in this study – here I will briefly describe the electrodynamic environment which causes it.

In the equatorial dayside F-region, a vertical upwards drift of plasma (both ions and electrons) across the magnetic field lines is driven by $E_\phi \times B_\theta$ action, a process generalised by Kivelson & Russell (1995). Here E_ϕ refers to the zonal electric field resulting from dynamo action at low and mid latitudes in both the E- and F-regions. The vertical drift lifts the plasma

to altitudes of around 800 km at the equator (Balan & Bailey, 1995b; Bailey et al., 1997). Under the action of pressure and gravity the plasma diffuses downwards and polewards along the magnetic field lines to form the crests at about $\pm 15^{\circ}$ dip latitude and 400 km altitude (Mitra, 1946; Huang, 1974). Despite higher loss rates at lower altitudes, the action of this 'plasma fountain' (Balan & Bailey, 1995a) is sufficient to develop the EIA, and the F-region maintains a higher plasma density at altitudes of 400 km than it does around 200 km (Rishbeth, 1988). Interaction of the E- and F-region dynamos leads to the LT distribution of E_{ϕ} (and hence, the forces acting to produce the EIA), and controls the structure of the mid-low latitude ionosphere at quiet times (Eccles, 2004). I now summarise the relevant parts of both dynamos.

3.3.1 E-region dynamo

The primary driver of the electrodynamics of both the E- and F-regions at non-polar latitudes is the insolation-powered neutral wind flow, **U**. Figure 3.2 shows the distribution of the horizontal component of **U** in the E-region. At this altitude, horizontal neutral winds are the result of upwards propagation of predominantly diurnal-mode tidal disturbances, resulting from solar heating of the troposphere and stratosphere (Tarpley, 1970; Kelley, 2009). In the E-region they give rise to hemisphere-scale vortices of horizontal current flow, causing the well-known Sq current system, a part of the E-region dynamo (Rishbeth, 1997; Heelis, 2004; Kelley, 2009).

In the mid-latitude contribution to the Sq system, interaction of polewards U_{θ} with B_r produces a current flow of ions in the westwards direction $(U_{\theta} \times B_r)$. This current occurs in the sunlit parts of both the northern and southern hemispheres, causing a build-up of positive charge at the dawn terminator with an associated increase in negative charge at the dusk terminator (Heelis, 2004; Kelley, 2009). To preserve global zero-divergence in the currents (Rishbeth, 1997), a zonal polarisation electric field E_{ϕ} results from this charge imbalance (this will modulate the '**E**' term in equation (3.1)). It is eastwards on the dayside and westwards at night, and maximal at the dip equator. The ionosphere is very highly conducting in the direction of the magnetic field, creating a 'short-circuit' along any magnetic field line and allowing electric fields generated in the E-region to 'map' to the F-region, and vice versa (Rishbeth,

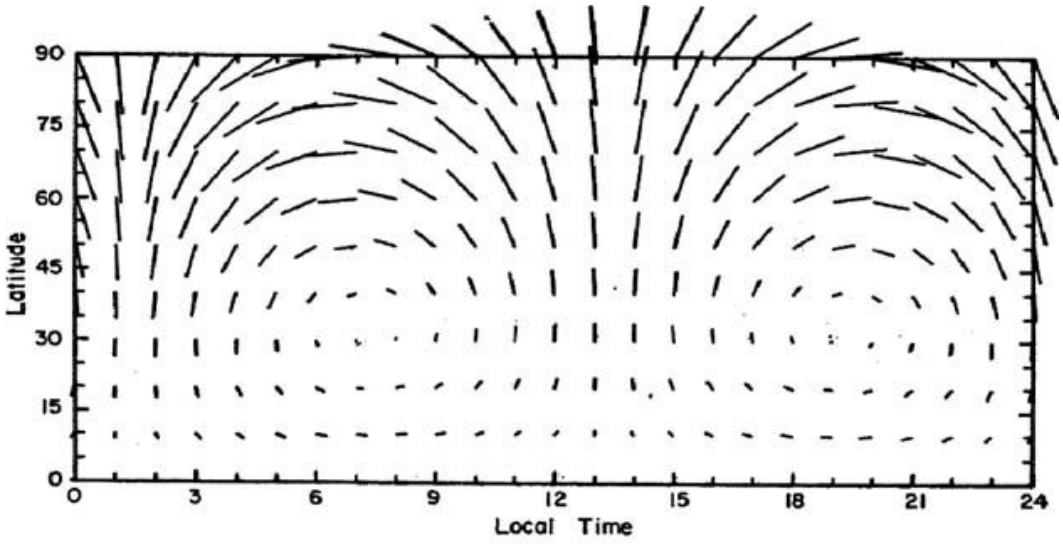

Figure 3.2: E-region horizontal neutral wind vectors in LT. The dominant solar-driven pattern of neutral winds at E-region altitudes, from Heelis (2004) and Tarpley (1970). Note the predominantly polewards flow at daytime mid-latitudes.

1988; Kelley, 2009). E_ϕ maps from E-region mid-latitudes to the equatorial F-region (Rishbeth, 1997). In this process the E_ϕ field remains zonal and (on dayside) eastwards and, by interaction with the horizontal magnetic field lines in the F-region, causes the $E_\phi \times B_\theta$ vertical plasma drift stated earlier. These interactions form the basis of the diurnal EIA cycle, and create some zonal current flow via the electric Lorentz force.

3.3.2 F-region dynamo and pre-reversal enhancement

Radar observations (*e.g.* from Jicamarca (Peru); Fejer et al., 2005) and models (Heelis et al., 1974) of the $E_\phi \times B_\theta$ vertical plasma drift trends show that their LT distribution can only be explained if they are modulated by the F-region dynamo (described by *e.g.* Rishbeth (1971), Rishbeth (1997) and Kelley (2009)). Like the E-region dynamo it is driven by insolation, yet the F-region neutral wind distribution $\mathbf{U}$ (horizontal components shown at equinox in Figure 3.3) is generated by in-situ absorption of solar radiation in the thermosphere. For ease of description I divide the F-region dynamo into its mid-latitude and equatorial parts.

The F-region dynamo at mid-latitudes operates on a similar principle to the E-region Sq system described above. Meridional polewards U_θ induces via $U_\theta \times B_r$ an east-west ion flow,

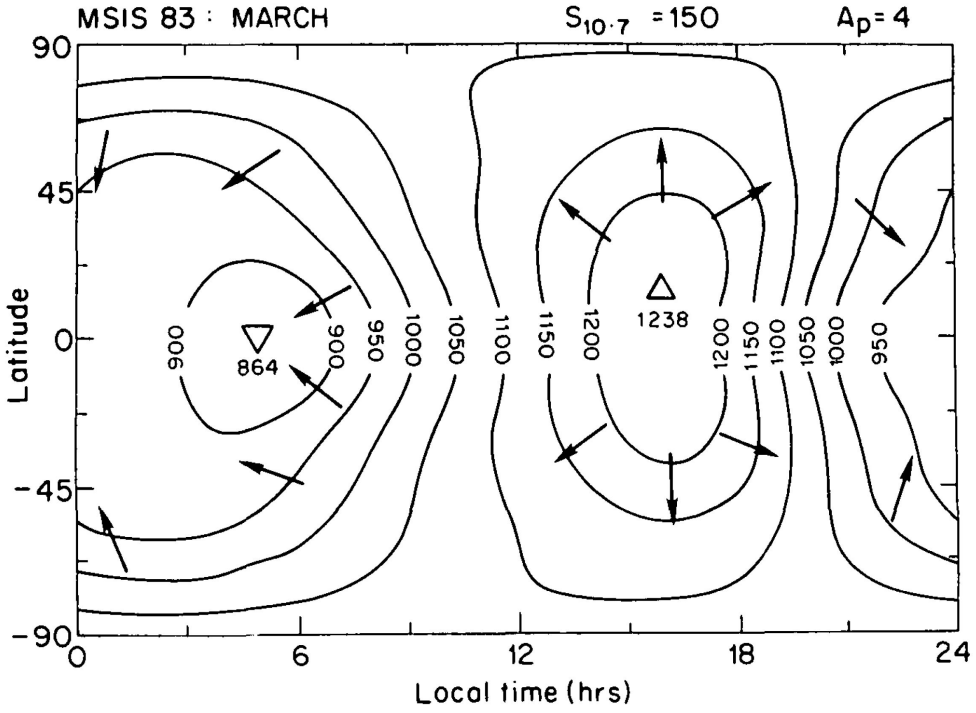

Figure 3.3: F-Region horizontal neutral wind vectors in LT. Thermospheric (F-region) neutral wind vectors, overlain with higher-altitude exospheric temperature contours (values in degrees K). From Rishbeth (1988). Note the offset in maximum temperature (hence also pressure) away from noon and towards dusk. This causes the predominantly westwards winds at the equator in daytime. The horizontal flow is shown – since gravity balances the vertical pressure gradient, neutral wind is considered constant with altitude in the F-region.

which leads to a charge build-up at the F-region solar terminators and an associated zonal polarisation electric field (Crain et al., 1993). The wind-induced current is stronger than the Pedersen current from the polarisation electric field (Forbes, 1981). Since the F-region is a weaker conductor than the sunlit E-region, the zonal current will be smaller than that produced at mid-latitudes in the E-region dayside, but will be generated on both dayside and nightside (subject to the latitude distribution of the plasma at F-region altitudes). At the dip equator, F-region zonal neutral winds U_ϕ flow westwards during daytime and eastwards during the night, causing (via $U_\phi \times B_\theta$) a vertical ion drift, whilst electrons are relatively unaffected by the wind forcing. With U_ϕ westwards and B_θ northwards in daytime, the resulting vertically downwards ion current at the equator causes a vertical charge imbalance, with negative charge at higher F-region altitudes. On the nightside the vertical and zonal directions are reversed, and positive charges accumulate at higher altitudes (as shown earlier in Figure 3.1). On the dayside

this charge imbalance is short-circuited via a meridional system of field-aligned currents flowing between the E- and F-regions. On the nightside the E-region conductivity is too low to permit this. To preserve global divergence-free conditions, a polarisation electric field (E_r) between the vertically separated charges at the equator is created in the nightside F-region which opposes the ion $U_\phi \times B_\theta$ drift (Rishbeth, 1997). The polarisation electric field balances the divergence resulting from the vertical ion drift (regardless of short-circuiting), though there may be some residual vertical current related to the Pedersen conductivity. As stated above, Hall conductivity is negligible in F-region so I am able to ignore E_r as a source of zonal current.

By itself, the equatorial part of the F-region dynamo generates no significant zonal current or zonal electric field, but the modulation of the E-region dynamo by the F-region does change the zonal electric field strength. This mechanism, concentrated at the dusk equator, is called the prereversal enhancement (PRE), described by Kelley (2009) and Heelis (2004). It is in essence a dusk-centred enhancement of the zonal electric field. Figure 3.4 schematically illustrates the PRE process. In daytime the polarisation electric field E_r in the equatorial F-region maps downwards and polewards along the magnetic field lines to become an equatorially-directed electric field E_θ in the E-region. E_θ drives a westwards Hall current J_ϕ at E-region low latitudes, the result of which is to intensify the distribution of negative charge at the E-region equatorial dusk terminator. In response, a localised enhancement (E_ϕ) in the E-region zonal electric field occurs on both day and nightside, each directed towards the dusk terminator. Mapped back to the F-region, the zonal electric field enhancement (E_ϕ) causes a sharp acceleration in first the upwards, then (to a lesser extent) the downwards, plasma drift velocity. Since plasma production rates do not change significantly over the occurrence of the PRE, this process acts to disperse the EIA plasma to higher altitudes, reducing the plasma density at lower F-region altitudes. The PRE's modulation of the LT distribution of vertical plasma drift velocity is illustrated by the effect on E_ϕ of including the impact of the E-F-region coupling, shown in Figure 3.5. The plasma drift trends form the EIA, and are a good proxy for the LT distribution of the zonal electric field strength, and thus for the electric field Lorentz contribution to the zonal current.

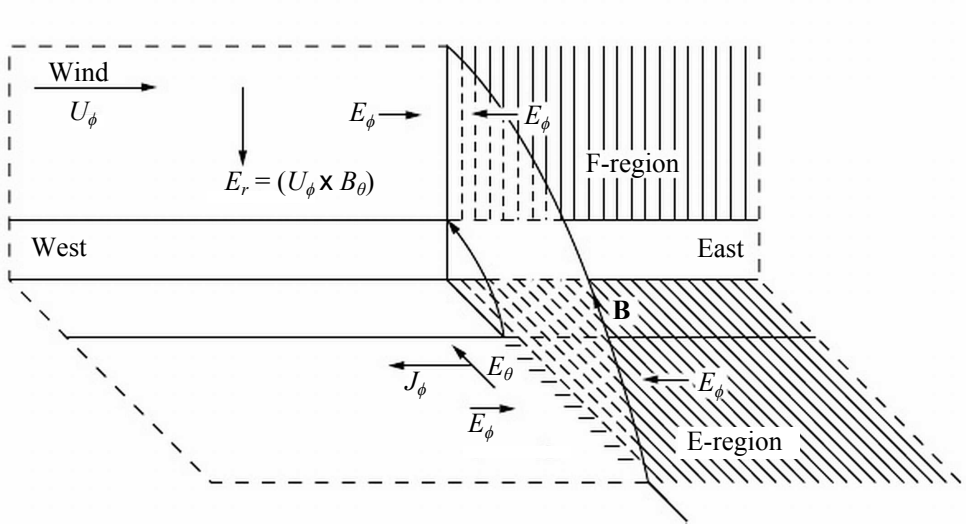

Figure 3.4: The prereversal enhancement mechanism (after Kelley, 2009, Figure 3.20). The field E_r 'maps' downwards and polewards (along the highly conducting magnetic field lines) into the E-region to become the equatorwards electric field E_θ. This field induces the westwards current J_ϕ. The polarisation electric field (E_ϕ), which appears to equalise the divergence (caused by J_ϕ), maps to the F-region.

The EIA plasma crests rise in the F-region throughout daytime (in accordance with the E_ϕ trends), with the peak crest density increasing in dip latitude to around 15°. At sunset the F-region plasma is dispersed and lifted to high altitudes by the PRE, where recombination rates are decreased (Kelley, 2009, page 99). Here the EIA (and the F-region in general) can persist until near dawn, despite the action of the reversed plasma fountain (Balan & Bailey, 1995b). After sunset, plasma production is greatly decreased, and the lack of upwelling plasma combined with the action of the plasma fountain intensifies the bifurcation of the EIA. The F-region eventually subsides and disappears almost completely before dawn.

The distribution of zonal current density at the altitudes of the Ørsted and CHAMP satellites in LT can be understood by the distribution of the zonal electric field strength, the dynamo action of the neutral winds, and the plasma density accumulations related to the EIA. The tendency for stronger zonal current near the dip equator is due to the latitude distribution of the EIA, increasing the Pedersen conductivity through collisional processes, as well as providing more plasma for gravity and pressure gradient forces to act upon. There are other less im-

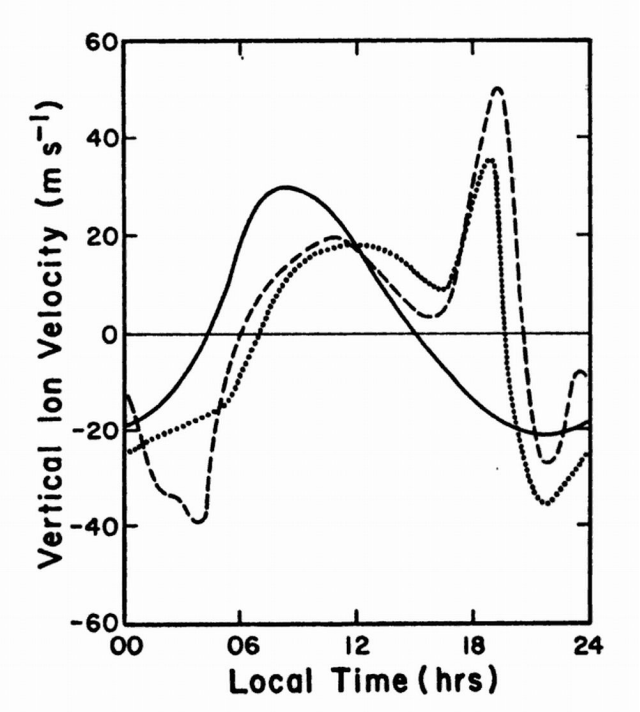

Figure 3.5: LT variation of vertical ion velocity at the equator (300 km altitude) from Heelis et al. (1974). The solid line is the tidal E-region electric field only. Dashed line is tidal E-region with F-region polarisation fields included. Dotted line is vertical drift velocity observed at Jicamarca around equinox (Aug.-Nov.) in 1968-1969. The effect of the pre-reversal enhancement is clear around 18 hours LT in the model curve which includes both E and F-region dynamo contributions.

portant sources of zonal current in the low latitude F-region, such as the electric polarisation fields produced by the divergence of the gravity and pressure gradient currents, and the effect of impinging electric fields from high latitude and magnetospheric processes. The relative importance of these effects at low latitudes is an open issue in ionospheric physics (Kelley, 2009).

3.4 Method details

3.4.1 Integral loop formation and data correction

Several efforts have been made to resolve electric current density estimates using magnetic data from satellites. Besides the recent studies mentioned in section 3.2, I also note the study of Suzuki & Fukushima (1982), who were the first to apply Ampère's integral to satellite magnetic data. To develop the method of resolving event-by-event in-situ current density estimates from satellite data, I draw on two recent approaches in particular. Stauning & Primdahl (2000) integrated signal from half-orbits of Ørsted to resolve cross-polar-cap current flow. They demonstrated a relatively small error arising from fictitiously closing the integral loop via the

equator. Ritter & Lühr (2006) described a method that could be used to resolve radial current density (projected onto the field-aligned direction) from the side-by-side configuration of the two lower-altitude satellites in the upcoming *Swarm* multi-satellite constellation mission of the European Space Agency (ESA). The basic process I use is an application of Ampère's law, which relates the satellite magnetic field measurements **B** to the current I flowing through a closed loop:

$$\oint \mathbf{B} \cdot \mathbf{dl} = \mu_0 I \tag{3.2}$$

where **dl** is a series of line elements which comprise the closed circuit and μ_0 is the permeability of free space ($4\pi.10^{-7}$ T m/A). In this case the closed circuit is a near-simultaneous overflight of Ørsted and CHAMP.

CHAMP and Ørsted have shared vector data coverage in the years 2000 to 2006. Within this period the mean LTs of each daily magnetic vector data file are compared for both Ørsted and CHAMP. Those which are within ± 2 hours of one another are retained (termed an epoch). The full set of epochs is shown in Figure 3.6 – each contains two LT sectors of orbital overlaps. These overlaps span the set of LTs twice, though the data coverage within these epochs is affected by data missing from Ørsted's ideal coverage. The impact of even short timespans of missing data is further exacerbated by the narrow UT (Universal Time) band in which the valid overlaps lie. Figure 3.7 shows the progression of the overlaps through the months of the year, and therefore the season, that is being sampled by each epoch.

For each epoch, the daily files typically number around 50. To reduce the time taken to process these data, the globe is partitioned into 4° squares (with 0.25° overlaps). In each partitioned region, each Ørsted data point is used to locate all 'nearby' CHAMP points, nearness indicated by $\pm 1^\circ$ of longitude and the time taken for CHAMP to complete half an orbit (± 0.65 hours). Duplicated CHAMP points are removed once all partitioned regions are processed, creating an archive of roughly-overlapping segments of Ørsted and CHAMP data. These segments are processed in turn. At the most closely spatially-overlapping point in each

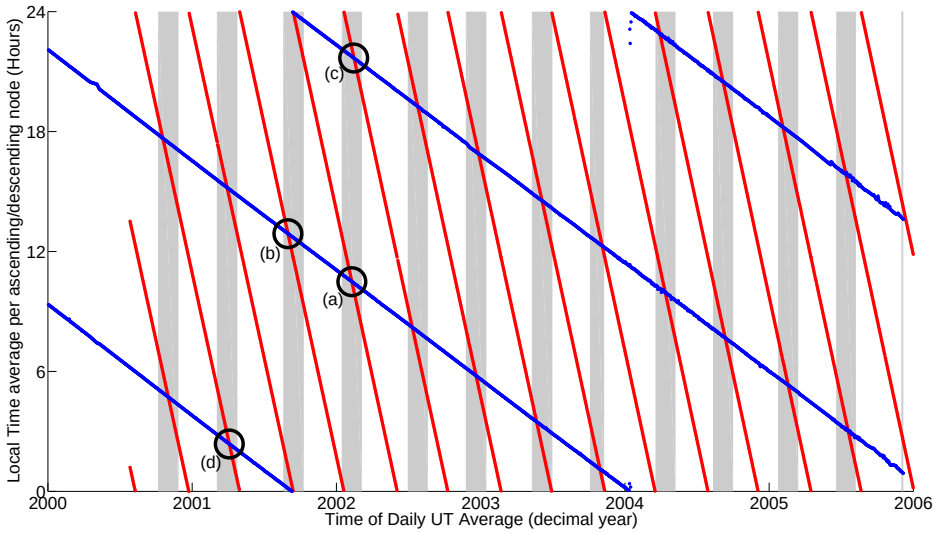

Figure 3.6: Available LT overlaps. Daily LT averages at mid-latitude for the ascending and descending nodes of the Ørsted (blue) and CHAMP (red) satellites. Shaded regions indicate ± 2 hours LT proximity of CHAMP to Ørsted. These data are interpolated: missing data affect the true overlap distribution. The black circles labelled (a) to (d) indicate the overlaps shown in the four panels of Figure 3.9.

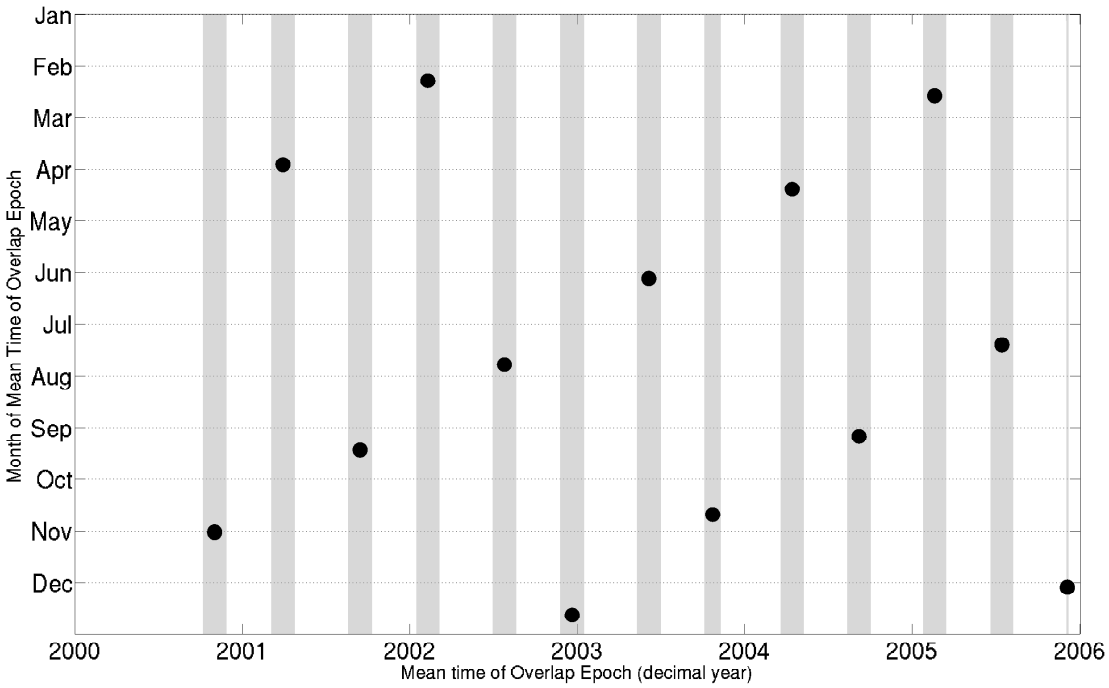

Figure 3.7: Progression of overlaps through the months of the year (hence, season). Shaded regions indicate ± 2 hours LT proximity of CHAMP to Ørsted, black points are the mean decimal year and season of the overlap epoch.

segment, I define contributing tracks of satellite data from Ørsted and CHAMP, each chosen to be an arc of length 2° in geographic colatitude (assuming no missing data), equal to the minimum radial spacing between the two orbits. Shorter arc-lengths do not increase the spatial or temporal resolution of the currents, and are proportionally more affected by missing data (based on tests not shown here). Ørsted and CHAMP have sampling intervals of 1.135 seconds and 1 second respectively, and each take around 30 seconds to cover the length of the arc. The integral loop is constructed from the two arcs, leading to around 70 data points in each loop.

The ECEF (Earth-Centered, Earth-Fixed) Cartesian coordinate system (hereafter 'Cartesian coordinates') is defined such that the x-axis points from the centre of the Earth towards the Greenwich meridian in the geographic equatorial plane. The z-axis points towards geographic North, and the y-axis completes the right-hand triad (*e.g.* Hapgood, 1992; Langel & Hinze, 1998). The **dl** in Equation (3.2) are formed in the integral loop from each successive pair of data points in the Cartesian frame as follows

$$\mathbf{dl} = (x_2 - x_1,\, y_2 - y_1,\, z_2 - z_1)\,. \tag{3.3}$$

Each line element defines the 'along-track' direction at that point in the integral loop. Since the relative flight directions of Ørsted and CHAMP reverse each epoch, I force the integral direction in each case to be north-to-south for Ørsted, and south-to-north for CHAMP, resulting in a consistent direction for the calculations, by which the convention for the subscripts 1 and 2 is defined. On the upper and lower parts of the loop, I form the line elements from the successive measurement points of the satellites. The radial connecting elements of the circuit have no data coverage between their endpoints, but still represent a part of the loop bridging two measurement points as I treat the radial elements in an identical manner to the line elements formed along the satellite tracks.

I do not apply data selection criteria prior to the calculation of the integral solution. From the magnetic data (**B**) in each pair of arcs the core ($\mathbf{B}_{core}$) and crustal ($\mathbf{B}_{crust}$) field estimates

are removed using the CHAOS-2 (Olsen et al., 2009) and CM4 (Sabaka et al., 2004) models respectively, producing the perturbation vector data $\Delta\mathbf{B}$ from

$$\Delta\mathbf{B} = \mathbf{B} - \mathbf{B}_{core} - \mathbf{B}_{crust}. \tag{3.4}$$

The effect of the crustal signal is not a significant factor in the results, due to its low amplitude at F-region altitudes. Tests (not shown) of the optimal spherical harmonic degree at which to apply the crustal correction concluded that degree 45 was adequate – the CM4 model was run at this setting. It is desirable to also remove the effect of the (highly time-variant) magnetospheric fields. However, available models of these fields are computed on at best an hourly basis, and the correction would have limited use inside the roughly 30-second timespan used in the overlap sets. An assessment of the error introduced by temporal lag within the integral loop is presented in section 3.6.1. The value of Ampère's integral (equation (3.2)) is dependent on the direction defined by the geometry of the satellites' overflight – this direction is expressed in the geographic (GEO) system (defined in section 3.3). To calculate the contribution of a single element of the integral loop (each successive pair of points) to the full circuital sum of equation (3.2), the magnitude of $\Delta\mathbf{B}$ in the direction of $\mathbf{dl}$ must be calculated, meaning that both vectors must be rotated into the along-track frame. Ritter & Lühr (2006) applied a similar process, treating each instantaneous 'quad' of measurement points as a locally horizontal surface. However, Ørsted and CHAMP both have slight (and differing) inclinations away from exactly-polar orbits, so the calculation region is a skew-quadrilateral (Weisstein, 2003) rather than a flat plane. As either Ørsted or CHAMP approaches the geographic pole, the inclination of the orbit results in a deviation between the geographic zonal direction and the strike-flight direction. The effect of this is to make the integral loops less zonally-facing (*i.e.* less parallel to a meridional line) at higher latitudes, and I have restricted my analyses to lower latitudes for this reason. Although this integral method may be applicable to high latitude data in order to resolve predominantly north-south currents, this would be significantly more difficult to apply and is beyond the scope of my study. The use of an along-track calculation frame for each pair of points around the integral circuit ensures that the coordinate frame in which the solution

is expressed is the one which best fits each overflight's geometry. In the following section I describe the process of rotating the magnetic vector into the along-track frame.

3.4.2 Rotation to the along-track coordinate frame

For the two magnetic vectors associated with each line element, I seek the component of each in the direction of integration. The conversion from GEO (the frame in which overlap instances are identified) to the along-track coordinate frame takes the form of three spherical rotations in the Cartesian frame. Therefore, I start this process with **dl** (equation 3.3) and $\Delta\mathbf{B}$ (equation 3.4) in the Cartesian frame (using the transformations described by *e.g.* Hapgood, 1992; Langel & Hinze, 1998). I use the Cartesian components (x, y, z) of **dl** to define the rotation process, which in turn is applied to $\Delta\mathbf{B}$. The along-track direction has been defined such that when **dl** is rotated to $\mathbf{dl}_{\mathrm{AT}}$ in the along-track frame, its x-component will point directly along-track. The transformation matrices for spherical rotations in the Cartesian frame about each of the principal axes are given here. For ease of reference, I will later refer to these in terms of the rotation angle and the principal axis as follows: a rotation about the x-axis by an angle of α is given as $\langle\alpha,X\rangle$.

I operate in a right-handed coordinate system (Boas, 2006) and the rotation matrices shown below will rotate in an anti-clockwise direction for a positive angle, if the viewpoint is on the positive part of the stationary axis, facing towards its origin. The rotation matrices for each of the three axes are:

$$\begin{aligned}
\langle\alpha,X\rangle &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{pmatrix} \\
\langle\alpha,Y\rangle &= \begin{pmatrix} \cos\alpha & 0 & -\sin\alpha \\ 0 & 1 & 0 \\ \sin\alpha & 0 & \cos\alpha \end{pmatrix} \\
\langle\alpha,Z\rangle &= \begin{pmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}
\end{aligned} \tag{3.5}$$

Rotation 1: rotate the x,y plane about z until the x'-axis points through the meridian of the first of the two points which make up **dl**. Calculation of the rotation angle α_1 is applied in the x,y plane, assuming conversion from radians to degrees, and correcting for quadrant as follows:

$$\alpha_1 = \begin{cases} \tan^{-1}\left(\frac{y_1}{x_1}\right) + 180^\circ & \text{if } x_1 < 0 \\ \tan^{-1}\left(\frac{y_1}{x_1}\right) & \text{otherwise} \end{cases} \tag{3.6}$$

The new coordinate system is calculated via

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \langle \alpha_1, Z \rangle \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{3.7}$$

Rotation 2: rotate the x',z' plane about y' until the x'',y'' plane is parallel to $\mathbf{dl}'$ and the first of the two points of $\mathbf{dl}'$ is at zero on the x''-axis:

$$\alpha_2 = \begin{cases} -\tan^{-1}\left(\frac{z'_2 - z'_1}{x'_2 - x'_1}\right) + 180^\circ & \text{if } x'_2 - x'_1 < 0 \\ -\tan^{-1}\left(\frac{z'_2 - z'_1}{x'_2 - x'_1}\right) & \text{otherwise} \end{cases} \tag{3.8}$$

Then

$$\begin{pmatrix} x'' \\ y'' \\ z'' \end{pmatrix} = \langle \alpha_2, Y \rangle \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}. \tag{3.9}$$

Rotation 3: rotate the x'',y'' plane about the z''-axis to put the x'''-axis in-line with both the points making up $\mathbf{dl}''$. If the second rotation was applied correctly, this angle should not have

to exceed 180° but I include the quadrant correction for completeness.

$$\alpha_3 = \begin{cases} \tan^{-1}\left(\dfrac{y_2'' - y_1''}{x_2'' - x_1''}\right) + 180^\circ & \text{if } x_2'' - x_1'' < 0 \\ \tan^{-1}\left(\dfrac{y_2'' - y_1''}{x_2'' - x_1''}\right) & \text{otherwise} \end{cases} \tag{3.10}$$

Finally

$$\begin{pmatrix} x''' \\ y''' \\ z''' \end{pmatrix} = \langle \alpha_3, Z \rangle \begin{pmatrix} x'' \\ y'' \\ z'' \end{pmatrix}. \tag{3.11}$$

$\mathbf{dl}_{\mathrm{AT}}$ is then $\mathbf{dl}'''$. The full rotation matrix R from GEO to the along-track frame is

$$R = \langle \alpha_3, Z \rangle \langle \alpha_2, Y \rangle \langle \alpha_1, Z \rangle. \tag{3.12}$$

The location coordinates must be rotated at each of the previous three stages, since the angle calculations require this intermediary information. However, R may be applied to the magnetic perturbation vectors in one step to derive

$$\begin{pmatrix} \Delta B_{\mathrm{AT}x} \\ \Delta B_{\mathrm{AT}y} \\ \Delta B_{\mathrm{AT}z} \end{pmatrix} = R \begin{pmatrix} \Delta B_x \\ \Delta B_y \\ \Delta B_z \end{pmatrix}. \tag{3.13}$$

The along-track component for each of the magnetic perturbation vectors is now their respective x-components. The mean magnetic perturbation vector in the integration direction ($\Delta\mathbf{B}_{\mathrm{AT}}$) for a single line element is then

$$\Delta\mathbf{B}_{\mathrm{AT}} = \left(\Delta\mathbf{B}_{2\mathrm{AT}} + \Delta\mathbf{B}_{1\mathrm{AT}}\right)/2. \tag{3.14}$$

The contribution of each line element to the integral current of equation (3.2) is the dot product of $\Delta\mathbf{B}_{\mathrm{AT}}$ and $\mathbf{dl}_{\mathrm{AT}}$. To obtain the current density from a summation of these contribution values, I must first calculate the area enclosed by the integral circuit – my approach to this is described in the next sub-section.

3.4.3 Calculation of the area enclosed in the Amperian loop

To calculate the area enclosed by the Ørsted and CHAMP arcs, I define a series of meshing triangles in the Cartesian frame, each comprising two locations from one arc (the base-line) and the third from the other arc. The base-line arc is switched for the next triangle in the mesh. The area A of each triangle is calculated via Heron's formula (Dunham, 1990)

$$\begin{aligned} sl_1 &= \sqrt{(bl_{1x}-bl_{2x})^2+(bl_{1y}-bl_{2y})^2+(bl_{1z}-bl_{2z})^2} \\ sl_2 &= \sqrt{(bl_{2x}-t_x)^2+(bl_{2y}-t_y)^2+(bl_{2z}-t_z)^2} \\ sl_3 &= \sqrt{(t_x-bl_{1x})^2+(t_y-bl_{1y})^2+(t_z-bl_{1z})^2} \end{aligned} \tag{3.15}$$

$$p = (sl_1+sl_2+sl_3)/2$$

$$A = \sqrt{p\cdot(p-sl_1)\cdot(p-sl_2)\cdot(p-sl_3)}$$

where sl_{1-3} are the triangle side lengths, bl indicates the two base-line points (with the subscripts 1 and 2 used as in section 3.4), t indicates the third point in the triangle and p is the semiperimeter of the triangle. The total area is the sum of the contributions A. This method is resilient to errors resulting from missing data since the base-line is a good approximation to a tangent for the satellite path, and remains so even if several points are missing. If either satellite arc has more points than the other arc, it is used as a series of base-lines with the third point provided by the end-point of the shorter arc until all the locations in the calculation region have been accounted for.

3.4.4 Along-orbit propagation of integral area and removal of invalid overlaps

I obtain the current density J flowing normal to the integral loop by dividing the value of the summed current contributions by the area A enclosed by the two satellite tracks, as follows

$$J = \frac{\sum_{i=1}^{n} \Delta \mathbf{B}_{\mathrm{AT}i} \cdot \mathbf{dl}_{\mathrm{AT}i}}{\mu_0 A}, \tag{3.16}$$

where n is the number of line elements in the integral circuit, equal to the number of data points.

Figure 3.8 shows a 24°-long arcspan of Ørsted and CHAMP data, with a typical integral calculation region (and its integration direction) superimposed. I find that these arc setups can be stably propagated along-orbit up to 12° colatitude in each direction from the crossover's centre-point, to increase data density. The along-orbit propagation involves incrementing the 2°-long Ørsted arc along-track by one Ørsted data point, and repeating the process of finding the best-fitting 2°-long CHAMP arc underneath it. The error introduced by the along-orbit propagation is discussed in section 3.6.1.

At this stage in the process, the temporal and spatial criteria used for selecting overflights are too accommodating, so the data set contains thousands of overflights which overlap in time but not closely in space, and vice versa. The set of along-orbit propagated overlaps are winnowed to a smaller set of viable solutions using a series of more stringent (but still principally geometry-based) rejection criteria. Table 3.2 shortlists and justifies the entire set of rejection criteria used to produce the results in this study. The error resulting from a given criterion being exceeded is small near the point chosen as the cut-off (rejection) value, though can increase greatly at extreme values of the criterion. The rejection values have been verified following tests (results not shown) with the synthetic *Swarm* data produced by Olsen et al. (2007). In these tests, the cut-off value for each criterion was chosen to be the point at which the increasing error in the current density estimates (with increasing values of the criterion) exceeded the

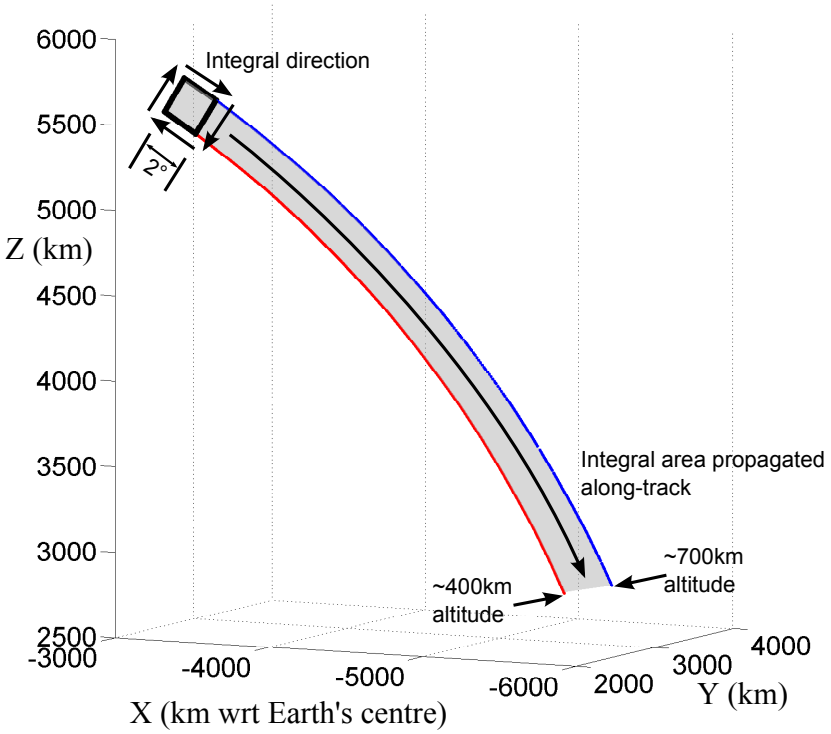

Figure 3.8: Cartesian coordinate geometry of overflight region and integral solution area setup. The Cartesian coordinate system is defined in section 3.4. Ørsted data locations shown as a part-orbit of blue dots, CHAMP as red. The black square indicates a circuit defined from an overflight of length 2° colatitude for each satellite. Note the lack of data on the radial connecting elements of the circuit. The arrows show the direction of integration, resulting in eastwards current flow being positive. The length of the longer arcs is the same as the distance the 2° setup will be migrated along-orbit, producing a solution at each Ørsted data point spacing. The shaded region indicates the area over which current is integrated.

amplitude of the 'scatter' present in the current density estimates at the zero-error value of the criterion. The most important of the rejection factors are based on the difference in the mean LT of the two arcs (their LT gap), the proximity to the magnetic poles (avoiding the auroral ovals), and any geometric aberrations caused by missing data (which could bias the calculations of integral surface area). Since the LT gap rejection criterion is set as the most stringent, it is uncommon for the full along-orbit migration length shown in Figure 3.8 to be used in the final results set.

After applying the rejection criteria, I obtain the final data set of overlaps for the epoch

Table 3.2: Invalid-solution rejection criteria

Parameter	Rejection Value	Justification
LT difference between arcs	> 0.096 hours (5.75 min)	$>$ 4x time-span of Ørsted arc
$\|$ Arc length $\|$	$> 10\%$ of 2° colatitude	N/A
Max $\|\mathbf{dl}\|$	$>$ max satellite altitude diff.	N/A
Min $\|\mathbf{dl}\|$	$\|\mathbf{dl}\| < 4$ km or $\|\mathbf{dl}\| > 8$ km	To identify geometrical errors
Loop surface area	outwith $\pm 15\%$ of expected value	Indicates area-selection error
Loop surface normal unit vector	θ component > 0.14	Empirical, greater at high latitudes
% missing points in integral loop	$> 15\%$	Empirical (tested with synthetic data)
Magnetic colatitude	within 40° of geomagnetic poles	Avoiding auroral regions
UT difference between arcs	> 0.096 hours	$>$ 4x trace-time of Ørsted arc
ϕ span between arcs	$> 1^\circ$	Empirical (tested with synthetic data)

being processed. Whilst the integration method discussed above can be used to obtain estimates of the average zonal electric current density in a three-dimensional loop defined between two satellites at different altitudes, the method requires certain simplifying assumptions. It is important to note that whilst the rejection criteria are as stringent as possible, I cannot correct for errors inherent to the method itself which result from these simplifications. For instance, although it is unlikely that the current is homogeneous between the two satellite altitudes, my method can only estimate the average current density and so is blind to departures from linear vertical trends in the currents. The bias introduced by this altitude difference is quantified in section 3.6.4, and leads to a consistent under-estimate of current density trends (in which between 20 and 40% of the signal is lost) by the method I have applied. However, the magnitudes of the trends I discuss in the next section are greater than this consistent loss of signal. The other errors inherent to the integral method, including the error introduced by the along-orbit propagation, are offset by careful selection and treatment of overlaps as discussed in section 3.6 and are also shown to have no detrimental effect on the results I present in the next section.

Further to the errors inherent to the integration method, the lack of any averaging in the representation of isolated LT sectors means that at any point, the current density estimates are subject to effects from the following controlling factors: solar activity, geomagnetic activity, season, longitude sector, colatitude and LT. These effects cannot be accounted for in the rejection criteria summarised in Table 3.2. The combination of the limited data coverage and the controlling factors precludes me from obtaining a coherent global picture of all the variations resolved in the currents. However, these drawbacks are acceptable given the purpose

of the method – to resolve only the currents affecting satellite magnetic data at the measurement altitude. A multi-satellite approach is preferable to the use of a single satellite in meeting these aims, since it obviates the need to parametrise (with inevitable error) the much stronger current contributions from above and below the satellites' altitudes, and allows event-by-event resolution of current density, removing the need to rely on a statistical approach. I assess the effect of each of the controlling factors in section 3.5.1 – my method is surprisingly resilient to controlling factors other than colatitude and LT. The full integration procedure described above is applied to every viable overlap within the 6-year timespan of mutually available vector data from Ørsted and CHAMP. In section 3.5.2, I use an ionospheric model to compare my estimates with a set of global 'typical' ionospheric background current density estimates.

3.5 Results

3.5.1 Local time sector analysis

Some systematic, global-scale features of the current density estimates are as follows. On the dayside the estimated current flow is strongest in the westwards direction, and has higher magnitude than on the nightside. In terms of latitude, the dayside current density is stronger near the dip equator than towards the poles (appearing to follow the EIA distribution). On the nightside, the current has no strong direction preference and exhibits weak latitudinal trends until after midnight, when the current density is strongest in the mid-latitudes. The magnitude of the current density estimates all lie within the range $\pm 0.1\ \mu\mathrm{A/m^2}$ (once solutions deemed to be invalid are rejected as per the criteria in Table 3.2). This is within the range reported by Ritter & Lühr (2006), though they focused on field-aligned currents at higher latitudes. Olsen (1997) used data from Magsat (altitude range between 350 and 550 km) to resolve meridional horizontal mean current density at dusk. Values in the range -30 to $+60\ \mathrm{nA/m^2}$ were obtained for the J_θ component between magnetic latitudes of $\pm 30^\circ$. At dusk, the Ørsted/CHAMP results for low-latitude zonal current density (J_ϕ) are in remarkably good agreement with the magnitude range of Olsen's Magsat values for J_θ, though typically occupy the lower magnitudes within this range. This difference is to be expected as the J_θ component will have a larger contribu-

tion from the inter-hemispheric field-aligned currents which permeate the F-region. Even if I were to measure J_θ, a difference in the results from Ørsted/CHAMP and Magsat would be expected due to the fact that the Magsat results are mean values (temporally smoothed), whilst the Ørsted/CHAMP results are instantaneous estimates. These have never been estimated at satellite altitude from multi-satellite magnetic data before and hence I do not expect a perfect match with previous estimates. Yet from the comparisons presented here, the Ørsted/CHAMP results appear reasonable.

I do not have current density estimates at all LTs as a continuous data set, since the usable data are recorded at times months apart, in different geomagnetic conditions and at different points in the solar cycle. However, the specific trends shown and discussed in this section all occur in more than one epoch and exhibit similar forms at different geomagnetic, seasonal and solar conditions. Thus, I consider the trends discussed here robust, and representative of the electrodynamics of the upper F-region ionosphere at the set of LTs shown.

Figure 3.9(a)-(d) shows electric current density estimates from four sectors of LT, each from a different series of overlaps as indicated in Figure 3.6. The sector in Figure 3.9(d) has a different range to the other subplots, but the figure uses a consistent vertical scale (all sectors span 0.16 $\mu\mathrm{A}/\mathrm{m}^2$). I have selected these particular epochs because they are among the 'quietest', geomagnetically speaking, and are the least affected by biases to the integral method (discussed in more detail in section 3.6). Each sub-plot of Figure 3.9 contains results spanning roughly 15 days in UT (non-continuous recording), in a band of LT approximately one hour wide. The approximate UT range (hence, season) and precise LT range are shown as text in each sub-plot. Note that gaps in the colatitude coverage indicate missing data from one or more satellites, and do not imply zero-values for the resolved current. I have used Quasi-Dipole latitude (Richmond, 1995) to keep the magnetic dip equator at 0°. Each data point is the result of applying Ampère's integral to a single Ørsted/CHAMP overflight of length 2° colatitude – eastwards current flow is positive, and westwards negative. The points are coloured according to the Dst (disturbance storm-time) index, a measure of global mid-to-low latitude geomagnetic

field activity. I include these sampled data to show their lack of effect on the trends I discuss – further examination of the effect of field activity is presented later in this subsection. Globally speaking, the current density estimates are most strongly organised in LT and colatitude (hence, they are assessed in this frame here). This indicates that the estimates respond most strongly to the plasma density accumulations of the EIA. The direction (sign) of the currents is then due to the relative dominance of the terms in equation (3.1) – I discuss this further in section 3.5.2.

Figure 3.9(a) covers the LT band 09:45 to 10:32 hours (hereafter the morning sector). The current density magnitude is largely invariant with latitude, though there are (rather indistinct) magnitude increases in both hemispheres from approximately 5 to 15° latitude (or the equivalent southern hemisphere latitude), which are likely related to the crests of the EIA. However, at this early LT, the EIA is not expected to have fully developed at CHAMP altitudes – this LT sector is included primarily to place the other sectors in better context. Since the morning sector shown here is recorded in northern hemisphere winter, I had anticipated the possible resolution of seasonal effects in the current density. Balan & Bailey (1995b) discuss seasonal effects on the EIA crest position, resulting from interhemispheric neutral wind flow in the thermosphere. Indeed, the magnitude of the relatively high-latitude current density estimates in this sector is increased in the northern hemisphere with respect to the southern hemisphere. However, this hemispheric difference in current magnitude is a non-seasonal effect, seen in each epoch covering the dayside. The hemispherical difference occurs in temporally adjacent epochs, between which the relative flight directions of the satellites will have reversed. Therefore I am able to rule out relative satellite flight direction as its cause – it is likely due to the unavoidable inclusion of data affected by significant (inter-hemispheric) field-aligned currents in the satellite overflight calculation region. I see no current density trends which are attributable solely to season, in any epoch.

Figure 3.9(b) covers the LT band 11:55 to 12:53 hours (hereafter the noon sector). Just two hours LT after the sector shown in Figure 3.9(a), the plasma environment has changed significantly. As ionospheric plasma continues to rise in altitude, the magnitude of the current density

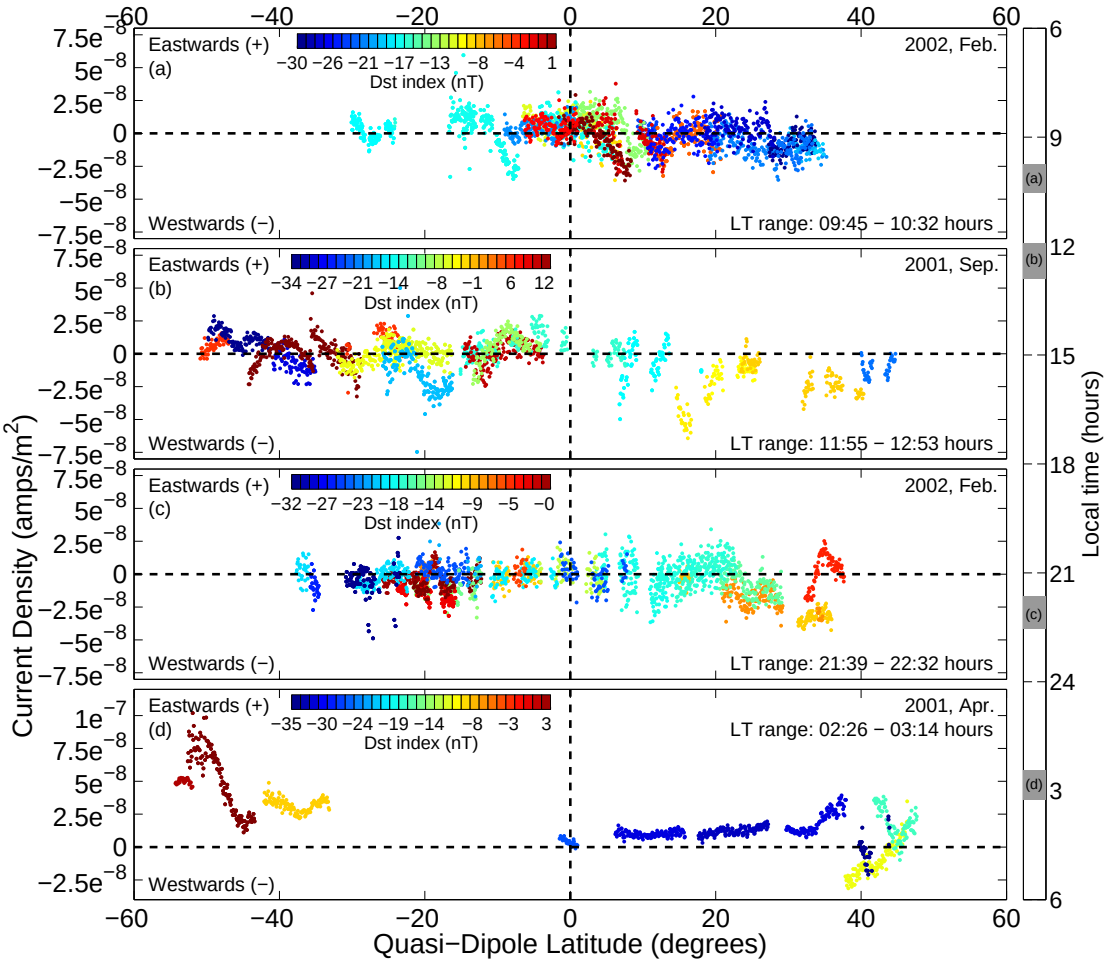

Figure 3.9: Solutions for zonal electric current density from the Ørsted/CHAMP overflights in four LT bands progressing from day through night: (a) 09:45 to 10:32 hours, (b) 11:55 to 12:53 hours, (c) 21:39 to 22:32 hours, (d) 02:26 to 03:14 hours. Each sub-plot contains results from about 15 days in UT (non-continuous recording), in a band of LT roughly one hour wide. Each data point is the integral-area-normalised result of applying Ampère's integral to a single overflight of length 2° colatitude. The points are coloured by samples of the Dst (disturbance storm-time) index. Eastwards current flow is positive; westwards negative.

estimates has increased, both near the equator and at higher latitudes. The changes at higher latitudes are obscured somewhat due to the limited latitude coverage of the sector shown in Figure 3.9(a). However, epochs (not shown) between the LT sectors in Figure 3.9(a) and (b) show a consistent increase in magnitudes. In the noon sector the bifurcated EIA crests between 10 to 20° latitude (and the equivalent negative latitudes in the southern hemisphere) are now clearer from the background scatter, indicating the development of the anomaly through LT.

Results from the LT range 21:39 to 22:32 hours (evening sector) are shown in Figure 3.9(c). Several electrodynamic processes have taken place since the noon sector shown in Figure 3.9(b). The EIA ascended to its full strength in late afternoon, continuing the trend of current density magnitudes increasing with LT throughout dayside. The PRE occurred about three hours (LT) prior to this snapshot. In this evening sector, plasma production has mostly ceased and the EIA has begun to subside, although the reversal will not have substantially decreased the altitude of the F-region yet. Despite the fact that the F-region is near its peak altitude in this evening LT sector, presumably increasing plasma volumes at LEO altitudes, the current density estimates here show lower magnitudes than much of the dayside (though the magnitudes are comparable to those of the morning sector). I consider this apparent discrepancy to be a combination of two factors. Firstly, Balan & Bailey (1995b) stated that the increased vertical plasma flow rate resulting from the PRE does not act to increase the amount of plasma in the F-region, rather dispersing it to higher altitudes. If the decrease in plasma density at CHAMP altitude is greater than the increase at Ørsted altitude, the resolved current density will decrease. Secondly, the fact that this evening sector has a degree of scatter more similar to the results in Figure 3.9(a) than Figure 3.9(d) indicates that a continued driving force is affecting the currents. Therefore the separate contributions to the total current density could be individually strong, but could act to cancel out when summed, producing no obvious resultant trend in the data.

As stated above I cannot resolve individual sources of current contribution, however, some insight is possible from theory. Maus & Lühr (2006) showed that a gravity-driven current at

LEO altitude encompasses this LT sector, having extrema at the EIA crest latitudes. The magnitude of the gravity-driven current in this sector is not greatly decreased from its dayside peak, corroborating the view that dayside-like magnitudes of current persist in the evening sector. The gravity current described by Maus & Lühr (2006) would cause wholly eastwards current. Yet in the estimates in Figure 3.9(c), the EIA crests (with apparent peaks near latitudes of $\pm 15^\circ$ QD latitude) show both westwards and eastwards current density increases. In addition to the pressure-gradient and Lorentz contributions (mostly westwards at this altitude) to the total current, Alken et al. (2011) discuss the effect of polarization electric fields (which create 'feedback' currents via the electric Lorentz force) on the currents originally responsible for the polarization imbalance. This effect will be discussed further in section 3.5.2. In summary, a combination of the plasma dispersal related to the PRE, and opposing sources of current, appear to define the current density estimates in this LT sector.

Figure 3.9(d) shows the LT range 02:26-03:14 hours (post-midnight sector). At this LT, F-region conductivity has decreased as the plasma descends into higher loss altitudes, and there is no direct electrodynamic driving force from the sun. This is apparent in the greatly reduced near-equator scatter in the results compared with the morning, noon and evening sectors. This nightside scatter reduction is typical across all epochs – where overlaps exist at the same UT and opposing LTs, the dayside values will always be 'noisier'. This suggests that the scatter seen on dayside is from an ionospheric rather than a magnetospheric origin. The low latitudes of this post-midnight sector are typified by lower magnitudes than on dayside. However, at higher latitudes (circa ± 30–50° magnetic latitude) there is a significant increase in current density. These mid-latitude intensifications occur in each epoch with data after midnight, exhibiting both east and west current flow with typical peak magnitudes of ± 0.05 to $0.1\ \mu\mathrm{A/m^2}$. This combination of latitude and LT is typically considered to be relatively free of currents, and the presence of these intensifications was not expected. The corresponding latitudes on dayside are free of this signal, with the exception of the afternoon sectors (not shown), which have typically strong values across a broad range of latitudes, but for which I have limited data coverage.

I suggest several possibilities for the source of the nightside mid-latitude intensifications. The high conductivity of the ionosphere in the direction of the magnetic field creates a 'short-circuit' along any magnetic field line, allowing electric fields to 'map' to higher or lower altitudes than the region in which they were generated (Rishbeth, 1988; Kelley, 2009). A number of high latitude/high altitude processes affect the mid-latitudes. The zonal current intensifications could be related to field-aligned currents from magnetospheric processes mapping down into the F-region (discussed by Rishbeth (1971)), or a result of the transmission of electric fields and neutral winds from the auroral region to lower latitudes as discussed by Heelis (2004). Schunk & Nagy (2009) discussed upflow of ionospheric plasma into the protonosphere (plasmasphere) during the daytime, and the subsequent descent of the plasma into mid-latitudes at night. However, the limited coverage of this epoch (and others in the same LT sector, not illustrated) depicts a signal in a narrow latitude band diminishing as the poles are approached, and latitudinal trends which do not correlate with Dst, implying a non-polar, low-altitude origin for the signal. The F-region dynamo (and associated polarization fields) is a possible candidate for the cause of the mid-latitude intensifications, but the intensification signal is not apparent south of 25° magnetic latitude in any nightside epoch – this could be explained by inter-hemispheric current flow along magnetic field lines. Assuming a dipole geometry, the altitude at the equator (apex altitude) of a specific field line is given by (Campbell, 2003)

$$a + p = (a + h)\sin^2(\theta_0) \tag{3.17}$$

where a is the radius of the reference sphere (here 6371.2 km), h is the apex altitude (above this reference surface), and p and θ_0 are respectively the altitude and colatitude of an arbitrary observation point. From equation (3.17), field lines which enter the Earth's atmosphere at 400 km altitude and 25° magnetic latitude have apex altitudes of approximately 1870 km. The upper boundary of the conducting F-region is somewhat diffuse, but it seems reasonable that field-aligned currents flowing between the two hemispheres in the ±25° magnetic latitude range could offset any local charge imbalance, reducing the horizontal current flow. As long as current can flow along these field lines the F-region is able to 'short-circuit' itself in a sim-

ilar manner to the coupling of the F-region dynamo and the E-region on dayside (discussed in section 3.3), preventing the build-up of large polarisation electric fields. Above 25° magnetic latitude, the apex altitudes are likely to exceed the conducting part of the F-region at this LT, preventing the (non-magnetospheric) magnetic connectivity of the two hemispheres. This would lead to a local charge build-up and an associated increase in the horizontal current density. The suppression of the intensification signal at higher latitudes is likely due to decreasing conductivity as plasma density decreases away from the EIA. The likelihood of a polewards conductivity decrease suggests that transmission of electric fields between the poles and the mid-latitudes would be difficult in this post-midnight LT sector.

Figure 3.9 presents a representative sample of my results, and analysis of the sample Dst values indicate that these estimated currents in these sectors are unaffected by the level of the Dst index. An analysis (not shown) of the entire set of current density estimates from all epochs alongside their Dst index values shows similarly little effect, even at extreme (high-negative) Dst. The sampled Dst in Figure 3.9 does not exceed 'medium' activity levels, but this range of values can still contain geomagnetic storm recovery periods (which can have low-negative Dst, but an energetic ionosphere and relatively high $dDst/dt$). Figures 3.10 to 3.13 each show a several-day span of the Dst index, overlain with the UTs (vertical red lines) of the crossovers from the sectors shown in Figure 3.9. Figures 3.10 to 3.12 show occasional instances of recording within a small storm recovery phase, but nothing which would be expected to significantly affect the results. Whilst Figure 3.13 shows recording times narrowly missing a large storm, the measurements taken in the later part of the storm's recovery phase do not affect the incidence of the nightside mid-latitude intensifications seen in Figure 3.9(d). The nightside mid-latitude intensifications are common to all epochs in that LT sector, most of which do not sample storm recovery phases (but which have poorer latitude coverage and are not shown here). I conclude that field activity (in the range sampled during the Ørsted-CHAMP overflights) has no discernible effect on my current density estimates.

In addition to the field activity and seasonal effects which I have discussed above, I stated

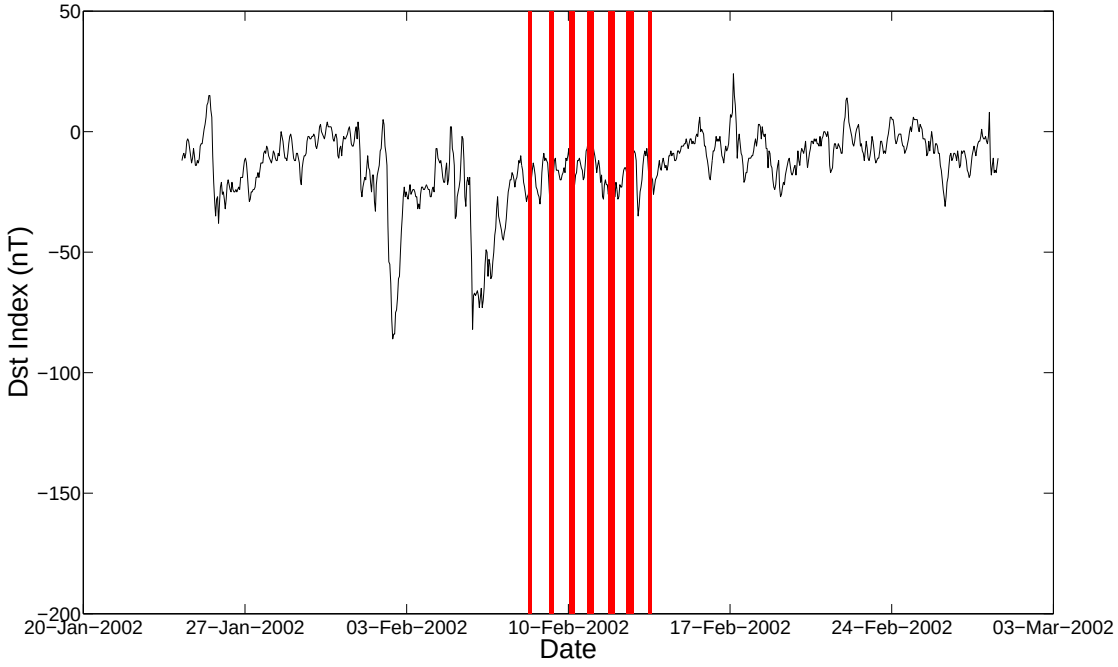

Figure 3.10: Black line is the hourly Dst index. The vertical red lines indicate the UTs of the crossovers in the dawn sector, shown in Figure 3.9(a).

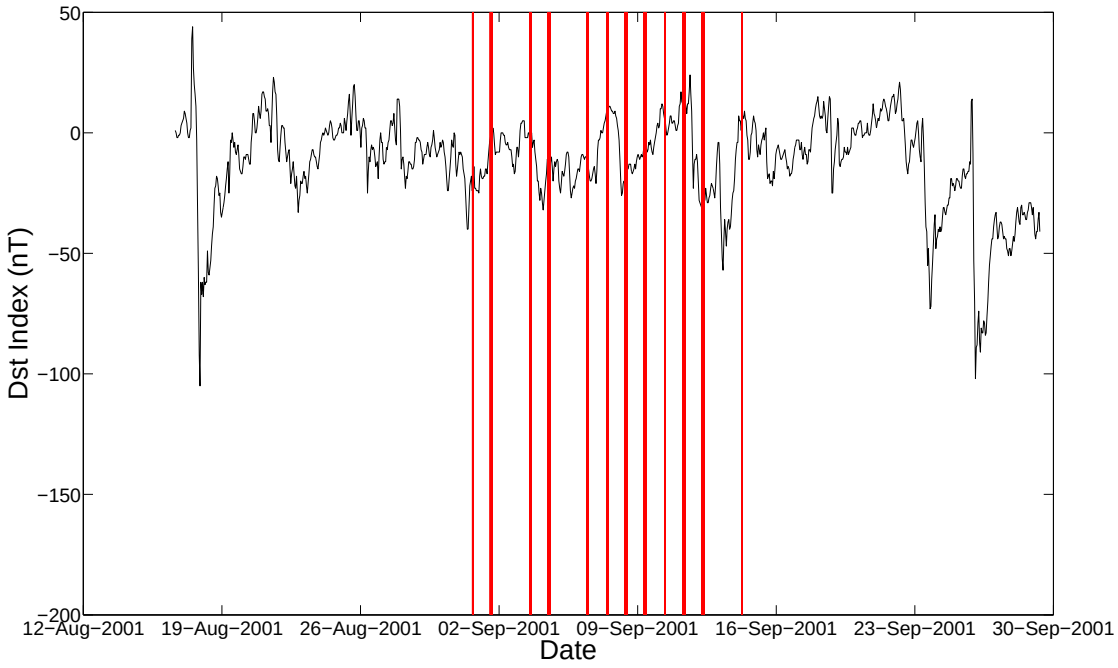

Figure 3.11: Black line is the hourly Dst index. The vertical red lines indicate the UTs of the crossovers in the morning sector, shown in Figure 3.9(b).

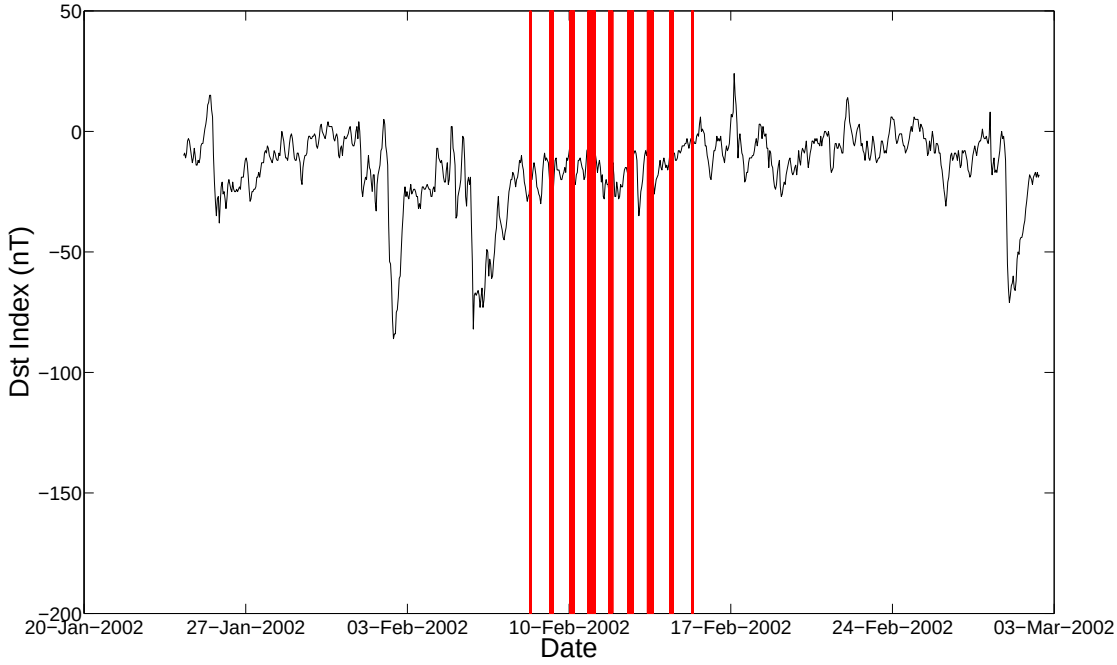

Figure 3.12: Black line is the hourly Dst index. The vertical red lines indicate the UTs of the crossovers in the evening sector, shown in Figure 3.9(c).

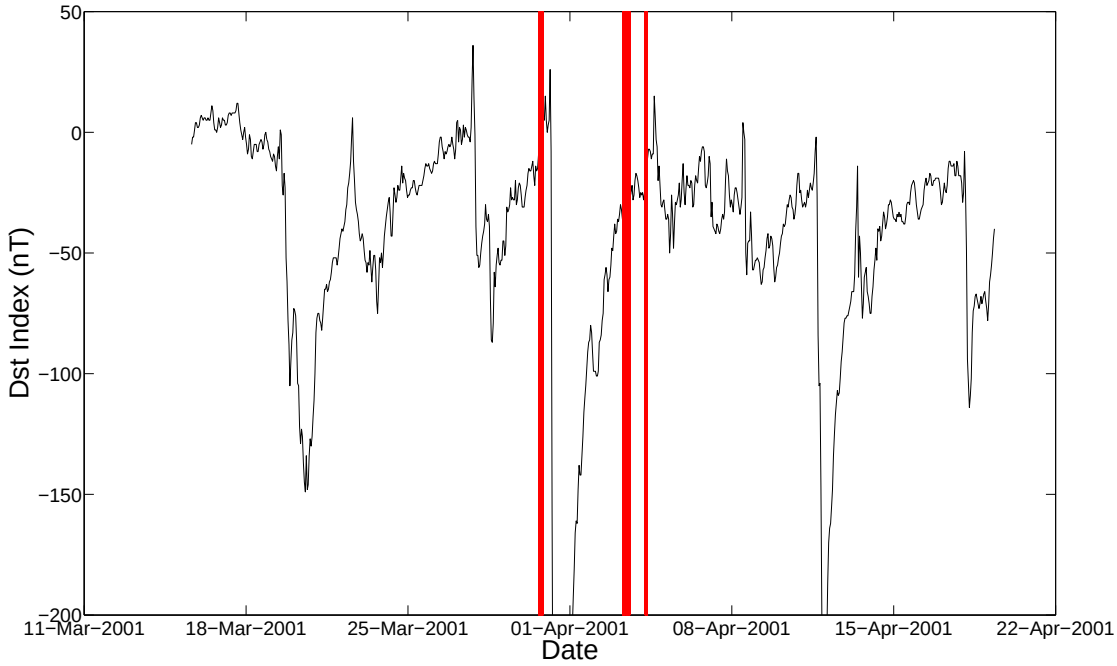

Figure 3.13: Black line is the hourly Dst index. The vertical red lines indicate the UTs of the crossovers in the post-midnight sector, shown in Figure 3.9(d).

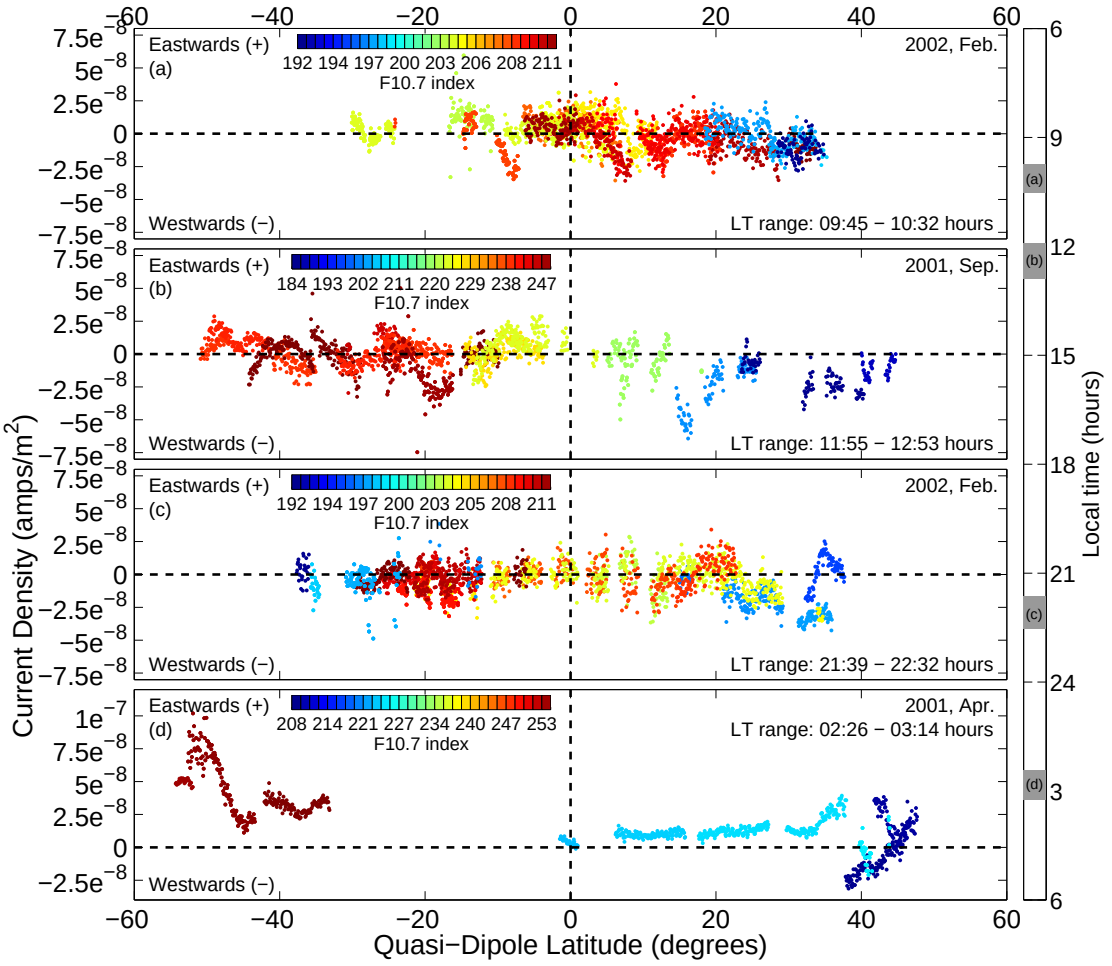

Figure 3.14: Identical to Figure 3.9, except that the points are coloured by samples of the $F_{10.7}$ solar flux density index (units of 10^{-22} W/m^2/Hz). Note that the solar flux density appears not to have any effect on the current density estimates.

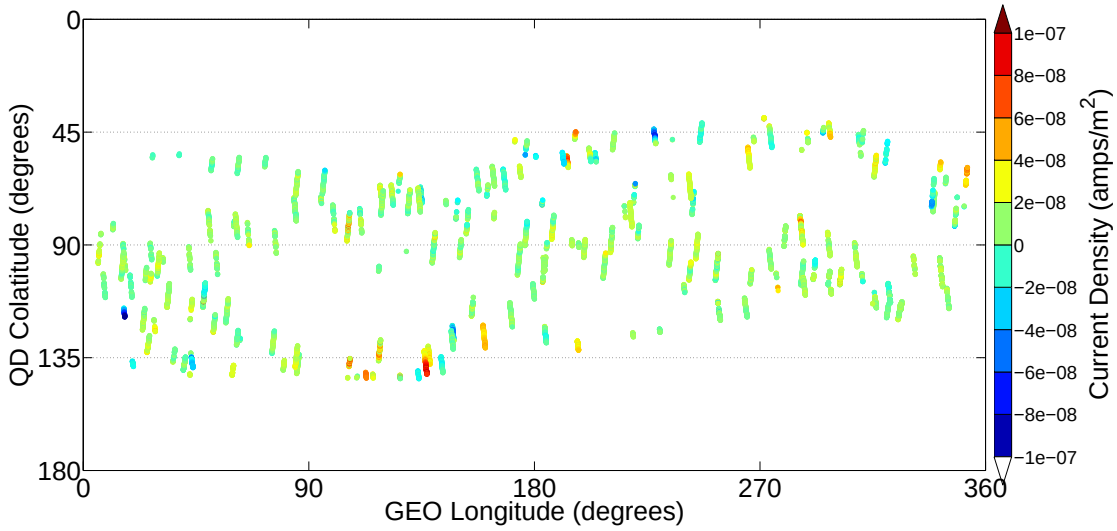

Figure 3.15: Current density results for the entire timespan in QD colatitude and GEO longitude coordinates, where QD indicates Quasi-Dipole coordinates (Richmond, 1995). These are the current density estimates from all Ørsted-CHAMP overlaps in the 6-year timespan of data. Here the solution value has been assigned to the centroid of each overlap area, with direction and magnitude according to the scale bar on the right (strong eastwards flow is red, strong westwards flow is blue). There do not appear to be any sustained longitude-sector trends.

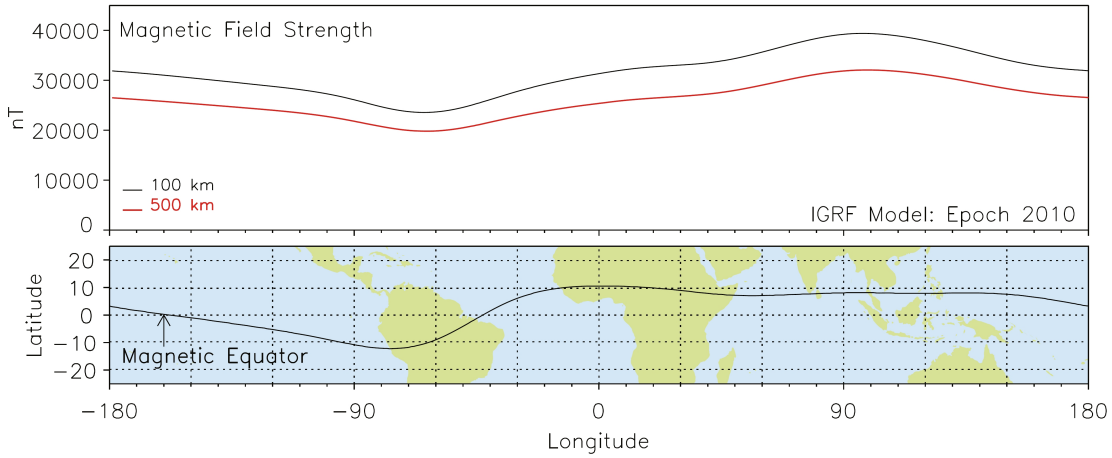

Figure 3.16: Lower panel: the geographic location of the magnetic dip equator. Upper panel: the difference in total field intensity along the dip equator (as given by the IGRF for 2010.0), at altitudes of 100 and 500 km. Comparison of this fluctuation with the longitude sector analysis in Figure 3.15, indicates that the current density estimates are not significantly affected by longitude sector. Image and description from Pfaff (2012).

in section 3.4.4 that my method is also susceptible to error from solar activity and longitude sector. Figure 3.14 shows current density estimates from the same overflights as Figure 3.9, but the points are coloured by samples of the $F_{10.7}$ index of solar flux density. The solar flux appears to have minimal effect on the current density estimates, despite the occasionally high values of solar flux sampled in the epochs discussed in this section. It is notable that Figure 3.14(d) has rather high $F_{10.7}$ values. There are two other epochs covering the same post-midnight LT sector (both of which exhibit similar mid-latitude intensifications, but which I do not show here). In the post-midnight sector, the epochs show the same trend in a variety of field conditions – I also note that the field activity level has no effect on the latitudinal pattern in the current density estimates, which is also a repeated trend between different epochs. As for the Dst index samples, an analysis (not shown) of the entire set of current density estimates from all epochs alongside their $F_{10.7}$ index values shows no effect on the current density estimates. This is rather surprising, given the reliance on the $F_{10.7}$ index (or a scaled value thereof) in modulating the parameterisation of ionospheric current systems in modern field models, such as CM4 (Sabaka et al., 2004). However, it is common (*e.g.* Sabaka et al., 2002) to empirically derive the proportionality factor used to translate $F_{10.7}$ values to ionospheric magnetic values, which should ensure an adequate fit if a representative set of data is used in the derivation of the scaling factor. This approach carries the assumption that the energy leading to neutral wind flow, and the energy that leads to enhanced conductivity, both impinge upon the ionosphere in the same spatial pattern (Kivelson & Russell, 1995). The result I present above, that $F_{10.7}$ has a varying aptitude in describing the controlling factors of different ionospheric current systems, would seem to go against this assumption. It would be instructive to compare $F_{10.7}$ with an index of currents measured at satellite altitude – the planned field-aligned current resolution with *Swarm* using the method of Ritter & Lühr (2006) would be a promising candidate. Lastly, Figure 3.15 shows the full set of current density estimates (from all epochs) plotted in a frame of QD colatitude and GEO longitude, to check for longitude sector effects. Figure 3.16 shows the longitudinal variation in magnetic field intensity, one of the possible factors which could cause a longitude sector dependence in the current density estimates. However, in Figure 3.15 there do not appear to be any systematic trends in longitude.

3.5.2 CTIP comparisons

I have assessed the likely electrodynamic fields behind the Ørsted/CHAMP estimates in each sector of LT and shown that, of the factors which can control the magnitude of my results (solar activity, geomagnetic activity, season and longitude sector), none impacts significantly on the trends I have discussed. However, I can only resolve the average current density and cannot distinguish between the separate contributing current sources in the satellite data. Hence, an independent validation of the broad-scale spatio-temporal trends in the Ørsted/CHAMP estimates is desirable. In this section I compare my results to predictions of the Lorentz, gravity and pressure gradient currents from the CTIP (Coupled Thermosphere-Ionosphere-Plasmasphere) model (Millward et al., 1996). CTIP is a self consistent, first-principles, 3-dimensional numerical model of ionospheric composition and temperatures. It is designed to highlight differences in small spatial-scale ionospheric dynamics between different geomagnetic activity conditions. Rather than relying on a parametrized electric field model for the low-latitude region, CTIP generates the low-latitude electric fields from the physics that determines the neutral winds, whilst the electron densities are determined from solar ionisation. It is possible that in this process the magnitudes of the electric fields and electron densities are being underestimated, and with them, the magnitude of the currents. By nature of its complete description of the relevant physics, I expect CTIP to provide a better spatio-temporal representation of the currents than a smoothed empirical model could, but do not expect an accurate match in terms of absolute values.

The default output of CTIP is a series of latitude-longitude grids spaced vertically in terms of pressure. When pressure is converted to altitude, this spatial coverage does not usually extend to the altitude of LEO satellites (dependent to an extent upon modelled solar activity). In this study I use the results from a special CTIP model run (Lühr et al., 2008a) in which the coverage was increased to 1000 km altitude. The extended data set is an hourly series of global 3D output from CTIP covering a single day in July 2003, in which the $F_{10.7}$ solar flux density is fixed at 130×10^{-22} W/m^2/Hz. The challenge of increasing CTIP's coverage up to 1000 km altitude made production of more than a single day's extended output prohibitively

difficult. In the following comparisons I use the CTIP predictions to describe the distribution of the contributing current sources for a 'quiet' ionosphere. This prevents the gravity- and pressure-gradient-driven contributions to the total current being overwhelmed by the Lorentz contribution. The trends discussed in section 3.5.1 are unresponsive to changes in Dst and solar flux, so a comparison with model values for a quiet ionosphere seems appropriate for identifying which current contributions are typically dominant in each LT sector (excepting periods of extreme geomagnetic or solar activity). The 3D grid spacing of the extended CTIP output is 2° latitude, 18° longitude and 50 km altitude (spanning 250 to 1000 km above a reference sphere of radius 6370 km). At each grid point the atmospheric parameters output by CTIP were input into equation (3.1), and current densities from the gravity, pressure gradient and Lorentz force contributions calculated. The first three spherical harmonic coefficients of the 2000 epoch in the IGRF-11 model (Finlay et al., 2010) were used to compute the tilted-dipole ambient magnetic field estimate used in this process as CTIP does not support a more complex parametrization of the magnetic field. $\mathbf{g}$ was calculated as in Lühr et al. (2008a, Appendix 1), and m_i assumed fixed as that of O+ (a reasonable assumption for F-region altitudes).

A plan view of the zonal current density according to CTIP at 450 km altitude is shown in Figure 3.17, with eastwards currents again positive. The locations of the Ørsted (white) and CHAMP (black) overflights (for several hundred overflight calculation regions) are overlain for the morning LT sector (as shown in Figure 3.9(a)). The vertical black dashed line indicates the location of the altitude-colatitude contour plot (Figure 3.18), which shows the zonal current density according to CTIP at 09:36 hours LT. The red box indicates a single overflight-calculation region for scale – the sides and top of this box are almost equal length in Cartesian coordinates. The altitude distribution of the overflights with regards to the EIA is clearly shown, though the anomaly has yet to fully bifurcate at this LT in the model. The Ørsted/CHAMP overflight locations vary both temporally and spatially, making sampling the CTIP values to the overflight locations difficult. I have applied a temporal simplification to the overflight locations, which occur in tightly clustered UT groups separated by a number of hours. The mean UT of each cluster of overflights is used in a 4D interpolation to compute a

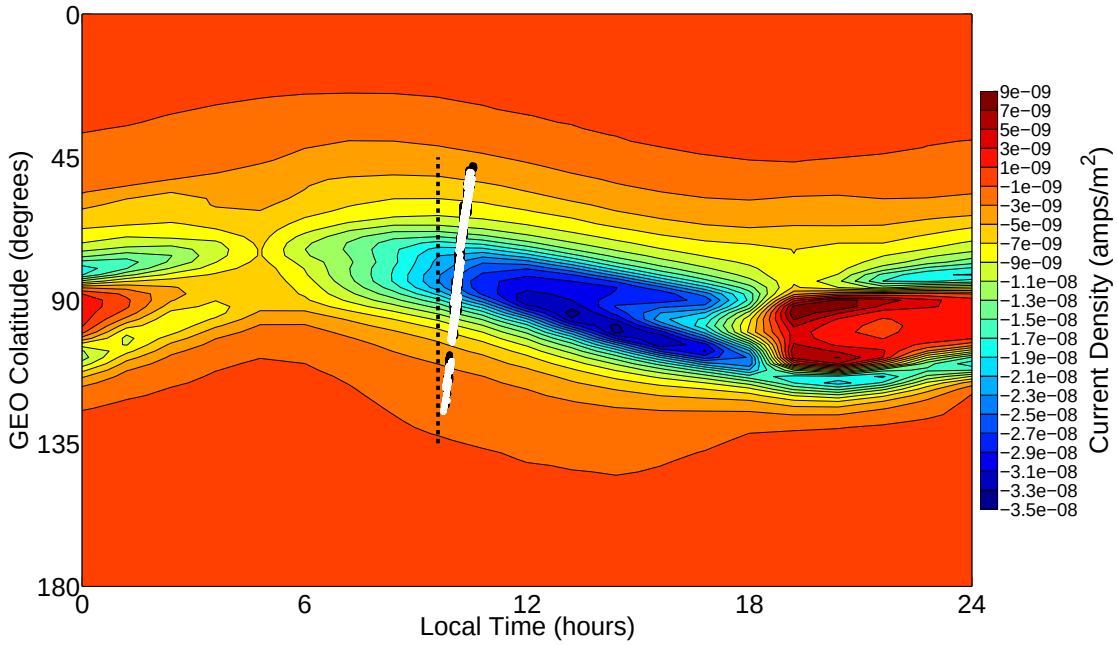

Figure 3.17: CTIP zonal current density prediction at a single altitude of 450 km in the LT frame. Eastwards flow is positive. The locations of the Ørsted (white) and CHAMP (black) overflights are overlain for the LT sector 09:45 to 10:32 hours. The vertical black dashed line indicates the location of the altitude-colatitude contour plot shown in Figure 3.18.

new global 3D grid of CTIP current density at the time of the overlap series. The new CTIP grids are each used in a tri-linear spatial interpolation to sample CTIP current density predictions to the locations of the Ørsted and CHAMP satellite tracks. These samples are averaged over the two altitude levels in each integral circuit, giving a single value per overflight for each force-contribution to the zonal current density prediction.

Figure 3.19 shows the same LT sectors as Figure 3.9, but with the Ørsted/CHAMP estimates smoothed (shown in black with a grey envelope of the non-smoothed estimates), and the CTIP predictions overlain. The values cover the same latitude more than once because each 1-hour-wide band of LT contains results from about 15 days in UT (non-continuous recording), and prior to smoothing, my estimates were grouped into clusters of similar UT. The blue points are the CTIP total zonal current density predictions, equal to the sum of the current contributions from gravity (green), pressure gradient (light blue) and Lorentz (magenta) forces. Note that the CTIP current densities have been (arbitrarily) multiplied by a uniform factor of 2 to aid

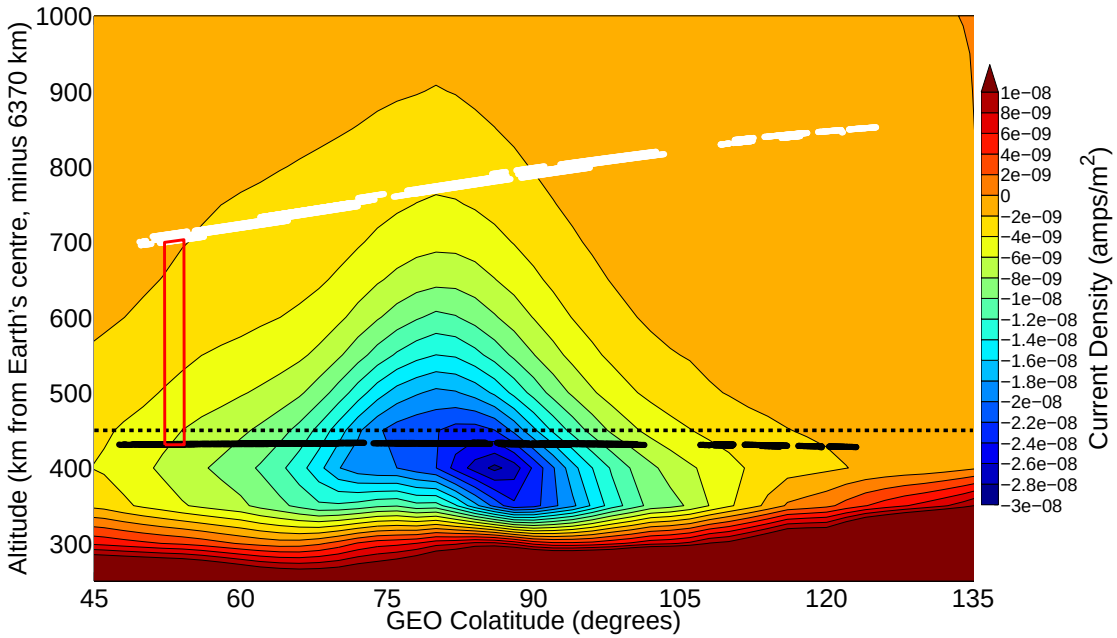

Figure 3.18: CTIP zonal current density prediction at a single LT of 09:36 hours for a July day. Eastwards flow is positive. The locations of several hundred Ørsted (white) and CHAMP (black) overflights are overlain for the LT sector 09:45 to 10:32 hours. The LT range of this sector is slightly outside of the LT of the CTIP prediction shown, but since 09:36 is one of the LT grid-nodes for CTIP, using this LT avoids the need for interpolation. The red box indicates a single overflight-calculation region for scale – the sides and top of this box are almost equal length in Cartesian coordinates. The black dashed line indicates the location of the colatitude-longitude contour plot shown in Figure 3.17.

in comparisons with the results from Ørsted and CHAMP. The CTIP prediction is typically an underestimate of the real-data current density estimate magnitudes, implying that CTIP is perhaps too effective at equalising local divergence in the ionospheric electrodynamics. Though the magnitudes of the current density estimates differ between CTIP and the satellites, the spatio-temporal agreement shows several key similarities. The CTIP model prediction resultant is dependent on the sum of several competing sources of current. Typically, where these sources diverge most strongly, the Ørsted/CHAMP estimates have their highest magnitudes. A notable exception to this is that the CTIP predictions in Figure 3.19(a) and (b) show the EIA in a non-bifurcated form, suggesting that the development of this structure is more rapid than is parametrized in the model. Furthermore, the CTIP prediction does not account for the aforementioned hemispherical difference in my current density estimates, and CTIP does not predict the nightside mid-latitude intensifications. Despite these differences, it is notable that

the magnitude offset between CTIP and the Ørsted/CHAMP estimates appears consistent between different LT sectors.

In Figure 3.19(a) there is a broadly similar latitudinal pattern to the magnitudes of the model predictions and Ørsted/CHAMP data estimates – a near-equatorial bulge with little extraneous signal at higher latitudes. However, the hemispherical difference in my current density estimates makes any more detailed comparison difficult, particularly at this early LT sector when the EIA has not fully developed. As noted above, the real data estimates appear to show a more bifurcated EIA than the CTIP prediction does.

Whilst in Figure 3.19(b) the CTIP predictions appear not to have increased in magnitude (from the previous LT sector) in-line with the Ørsted/CHAMP data estimates, this is likely due to a decreased plasma pressure gradient contribution and an increased gravity current contribution, affecting the resultant total magnitude of the prediction. The latitude of the CTIP peak magnitude in this sector is in the southern hemisphere since the CTIP prediction is for northern hemisphere summer, whilst my estimates in this sector are for the equinox. Note the magnitude peak in the Ørsted/CHAMP data estimates between 5 and 20° QD latitude – this is thought to be due to the EIA crest. If this is true, then regardless of the season of the CTIP prediction, the model again appears to have insufficient EIA bifurcation in this LT sector.

In Figure 3.19(c) the CTIP values show a magnitude decrease in-line with the decrease in the Ørsted/CHAMP data estimates from dayside LT sectors. In this evening sector the EIA crest positions in the data estimates and the model prediction appear in much better agreement than on the dayside, though this is less clear in the northern hemisphere. The cause of the northern hemisphere intensifications in the Ørsted/CHAMP data estimates at higher latitudes is unclear – they could be related to bias from field-aligned currents, or could be related to the mid-latitude intensifications seen more clearly in the post-midnight sector. CTIP does not predict them, for which I offer reasons in the next sector's analysis.

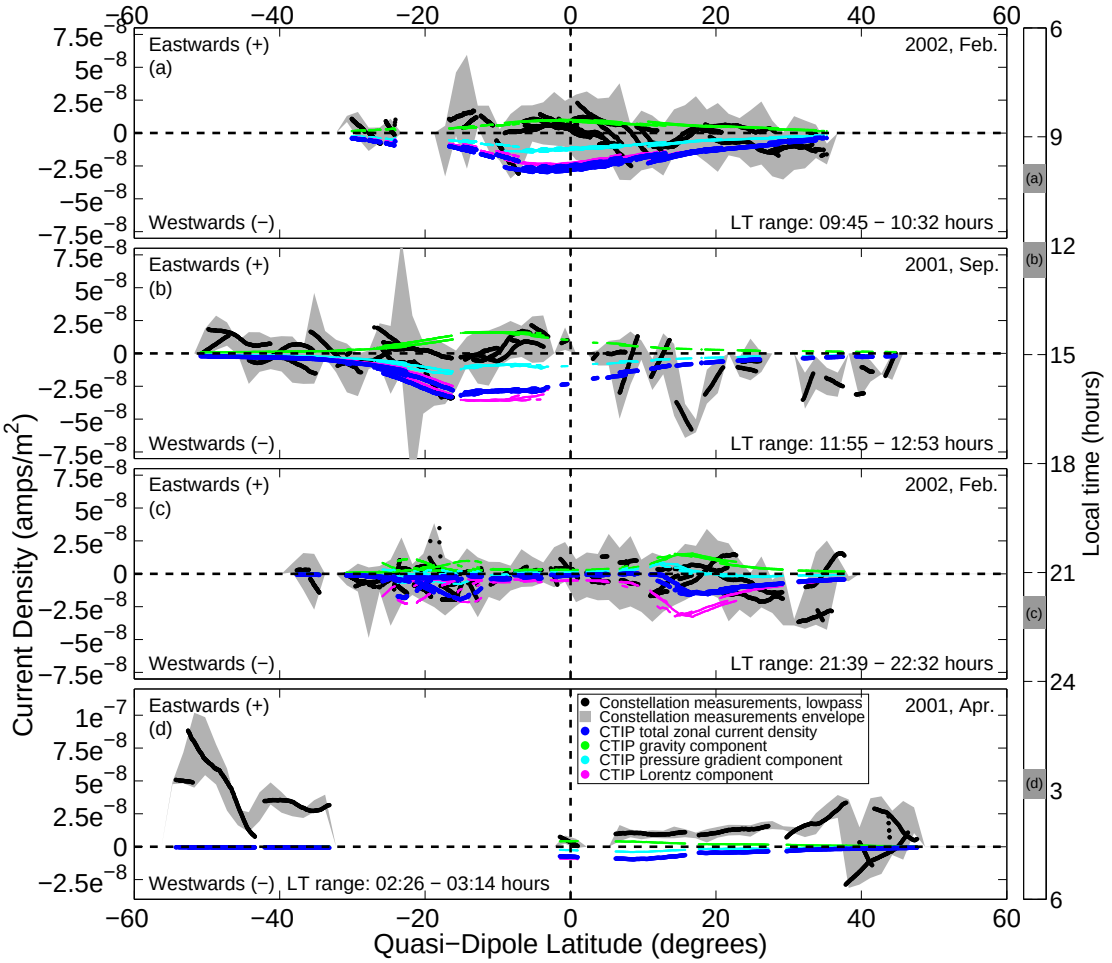

Figure 3.19: Comparison of CTIP predictions and Ørsted/CHAMP overflight estimates of current density. The subplots cover the same sectors of LT as Figure 3.9. The dark blue points are the CTIP predictions for total zonal current density sampled to the Ørsted and CHAMP locations, then averaged between the two different altitudes and multiplied by a constant factor of 2 to aid in comparisons with the satellite estimates. The modelled contributions of each force to the total zonal current density are shown as follows: green, gravity; light blue, pressure-gradient; magenta, Lorentz. The sum of these contributions gives the dark blue points. The black points are the satellite-data estimates from Figure 3.9, smoothed with a 50-point lowpass filter to remove the scatter and expose the underlying trends to aid comparison. The grey shading is a 1.5° latitude-smoothed envelope of the satellite estimates. The envelope was smoothed in the latitude coordinate frame, the black points in the UT frame. Eastwards current flow is positive; westwards negative.

In Figure 3.19(d) both data and model show a marked decrease in near-equatorial magnitudes, though the agreement worsens at higher latitudes. Without direct ionising irradiation from the Sun, the nightside F-region is subject to a different balance of controlling forces than the dayside, and can exhibit mesoscale structure which would not be dominant during sunlit hours (Kelley, 2009). Although the dayside is host to stronger currents with an associated higher error in their prediction, the nightside ionosphere is considered more challenging to parametrize effectively than the dayside because the dominant forcings are less easy to predict. CTIP has evidently not parametrized the mid-latitude intensification signal – I suggested earlier that these intensifications could be due to the inability of the mid-latitudes to connect via field-aligned currents at this LT. A parametrization simplification adopted by CTIP, intended to force the global mean divergence in the currents to zero, closes the horizontal currents via field-aligned currents. It could be that this simplification is preventing the appearance of the mid-latitude intensifications in CTIP's prediction.

The CTIP comparisons use a single day of data prediction made with fixed values for solar activity and season (representing northern hemisphere summer, moderate solar flux and quiet geomagnetic conditions). At higher geomagnetic activity levels, the contributions from the gravity and pressure gradient forces will cease to affect the Lorentz contribution significantly. The balance of contribution values to CTIP's total zonal current density shown here will not be applicable in all conditions. Despite this simplification, the comparisons have been instructive in assessing the input of each current type into the bulk signal typical for a 'quiet' ionosphere, as well as validating the spatio-temporal trends in my results.

3.6 Assessment of sources of error and uncertainty

Whilst my results originate from individual crossovers of the satellites, they are not direct measurements. Compared to the magnitude of the full geomagnetic field, I am isolating a relatively small signal from much larger background trends. A discussion of possible sources of error follows. Figure 3.20 shows the distribution of the estimates in LT. The values highlighted in the five boxes will be discussed in the following sections.

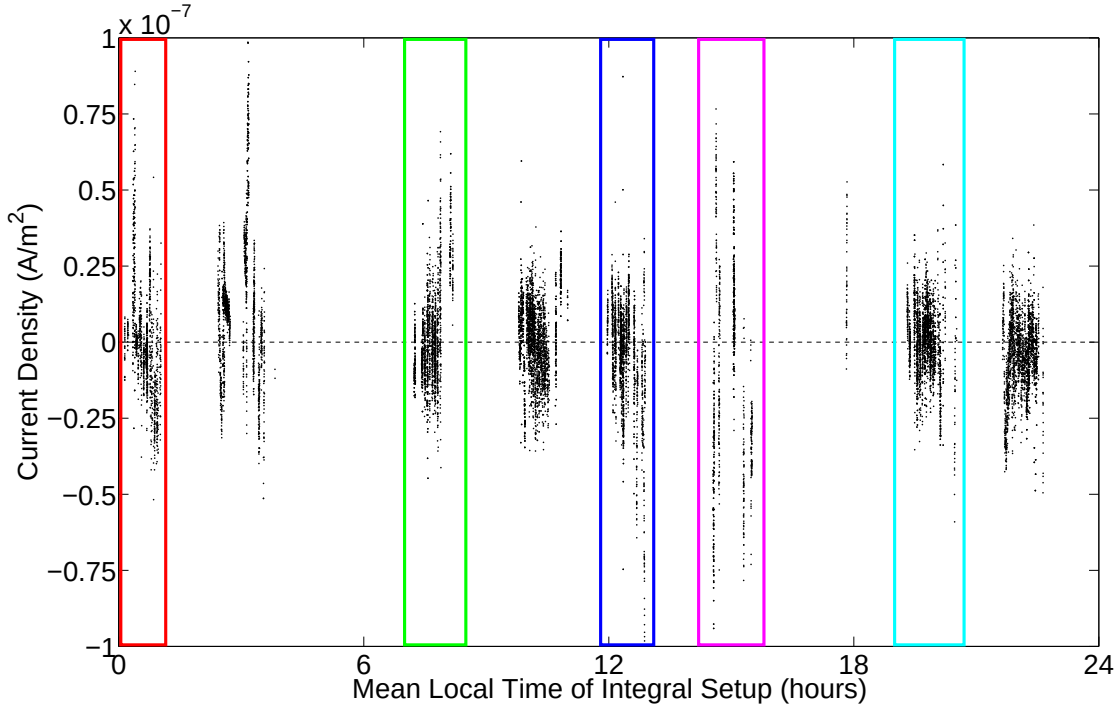

Figure 3.20: Current density versus distribution of Ørsted and CHAMP overlaps for all LTs. The EIA-related trends here are scattered due to the interference of several dependencies in the results. Here I highlight the 'boxed' sets of values: the red and dark blue boxes show a pair of current density estimates from data recorded at the same UT, and the light blue/green boxes show a different set of results from an identical UT (though this is different to the UT of the other boxed values). These will be discussed later, in relation to Figures 3.24 to 3.27. The values in the magenta box will also be discussed below, in relation to Figure 3.22.

3.6.1 Temporal lag permitted in integral loop

Part of the appeal of studying low latitudes is that the response times of the current distributions to electrodynamic forcing are large compared to the time taken to complete an overflight. For this reason I have not allowed for changing electric flux in equation (3.2) (doing so would make this the Ampère-Maxwell law). However, the signals from the magnetospheric magnetic fields are highly time-variant. The overflights are near-simultaneous, but not precisely so. The propagation along-orbit increases this difference, especially when the satellites' orbits are in different directions. Any temporal offset between the two contributing tracks in the integral loop will have associated different measurements of the magnetospheric fields. This will act as a source of noise in the data. Here I assess the severity of this effect on the current density estimates.

The mean time taken by Ørsted and CHAMP to trace a 2° long arc is roughly 30 seconds. The results shown in Figure 3.20 (and elsewhere throughout this chapter unless stated otherwise) have been allowed a time difference between the mean LTs of each arc of up to ± 5.75 minutes (likewise for the mean-UT difference). Therefore, the time difference between the arcs can be more than 5 minutes greater than the time taken to record the data for a single arc. The temporal lag allowed between the arcs is generous in order to increase data coverage. Here I examine the effect of changing the time difference cut-off.

The trend for variation in solution value with increasing LT gap is shown in Figure 3.21, which spans ± 5.75 minutes (0.096 hours) on the abscissa. I point to the lack of difference in lateral trend in this chart as good evidence that the calculation method is temporally steady, for as far as can be resolved in time without incurring overflight-geometry-based errors in the solutions. However, this does not account for LT-dependent effects. In Figure 3.22, I show the data from Figure 3.20 with the allowed time gap reduced to just over one minute (the ± 0.02 hours span in Figure 3.21). Note that the magnitude of the boxed values does not differ significantly from Figure 3.20. I surmise that the temporal error imposed by the choice of arc length does not significantly affect the solutions.

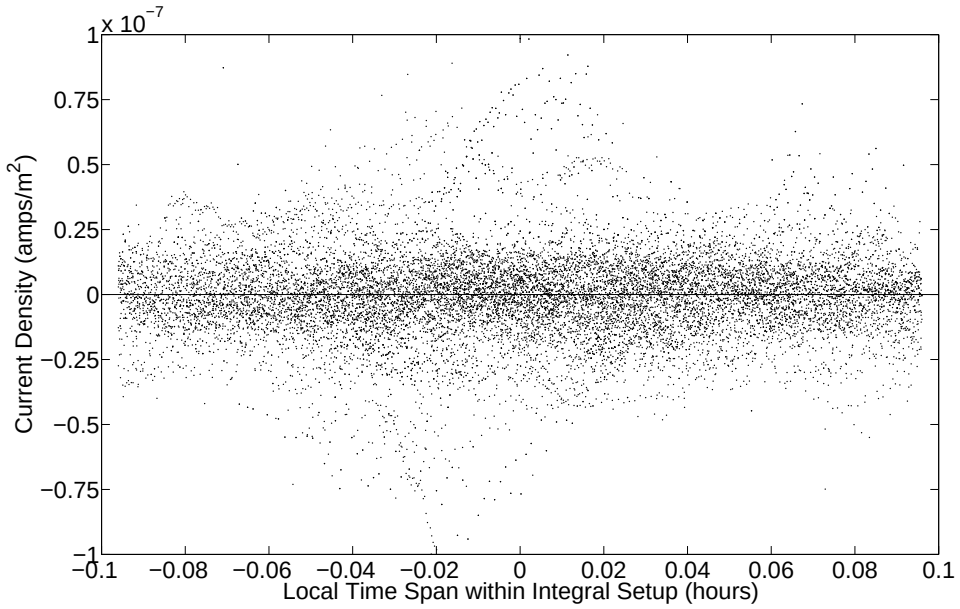

Figure 3.21: LT difference within calculation region. Each point is the mean LT of the Ørsted arc minus that of the CHAMP arc. The x-axis spans the time interval ($\pm$5.75 minutes) between LTs allowed in the calculations shown in Figure 3.20. The scatter is laterally uniform across the chart, indicating that the calculation method I use produces equally reliable results throughout the time taken to record a full loop of data from the two satellites.

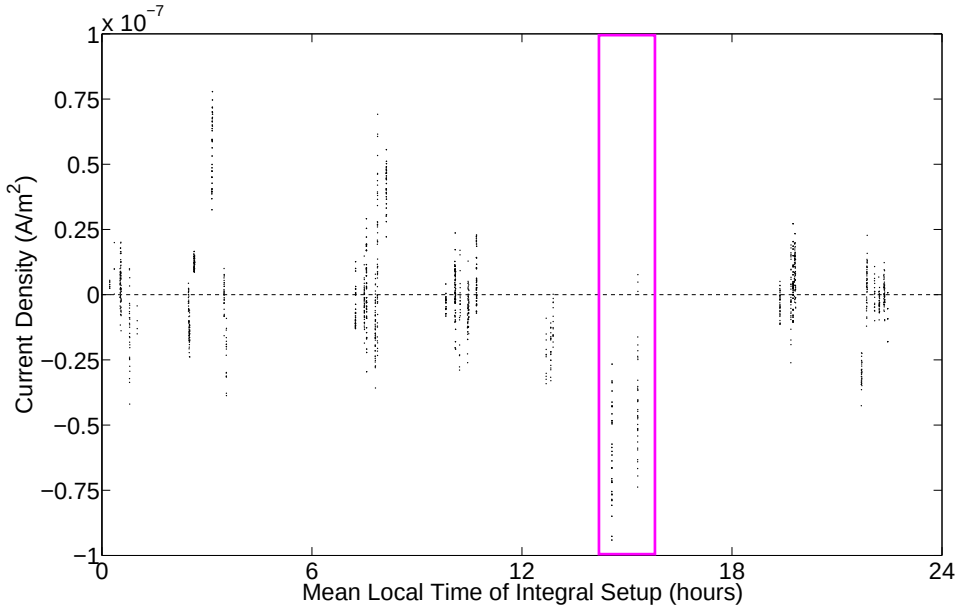

Figure 3.22: Current density versus distribution of Ørsted and CHAMP overlaps for all LTs. Identical to Figure 3.20, except that the allowed mean time-gap (in LT) between the contributing arcs has been decreased from 0.096 hours (5.76 minutes) to 0.02 hours (1.2 minutes). Note that the amplitude of one of the the highest-magnitude sectors (boxed) is not diminished by this step.

3.6.2 Effect of attitude error

The rotation of satellite magnetic data from the frame of the vector fluxgate magnetometer (VFM) to an Earth-centred, Earth-fixed (ECEF) frame involves an in-flight estimation of the three Euler rotation angles, which specifically are used to rotate the magnetic vector from the magnetometer to the star tracker frame, with the subsequent rotation to the ECEF frame determined from the star-tracker orientation and satellite position, as discussed by (*e.g.*) Olsen et al. (2007). The Euler angles are estimated under the assumption that the currents described in this chapter do not exist. It is possible that the rotation process could screen out the signal I am attempting to resolve. Worse, the results of the integral could simply be due to attitude offsets between the Ørsted and CHAMP satellites. Here I assess the likelihood of this occurring.

Simulated data from the ESA's upcoming mission *Swarm* – described in the End-to-End mission simulator study (Olsen et al., 2007) – were used to calculate Ampère's integral in an identical manner to the actual data from the Ørsted and CHAMP satellites. Since the simulated *Swarm* data are all based on potential fields, I expect zero current density except for errors introduced by time differences within the integral setup. The blue points in Figure 3.23 show the *Swarm* integral solutions, with an arbitrary vertical offset applied to aid comparison with the following data series.

The Ørsted and CHAMP Euler rotations are determined to better than 20 arc seconds, equivalent to a combined magnetic error of around 5 nT in a 50,000 nT ambient field. The red points in Figure 3.23 show that the result of directly applying a 10 arc second attitude error to the first Euler angle (for the lower satellite only) used in the VFM-ECEF rotation is minimal – this is also the case for 10 arc second errors applied to the second and third Euler angles (results not shown). Note that the red points have been vertically shifted in an opposite manner to the blue points for clarity.

Testing this effect further, I applied an extreme attitude error to the first Euler angle, changing the rotation by 431 arc seconds, equivalent to 25 % of the total rotation amount. The results,

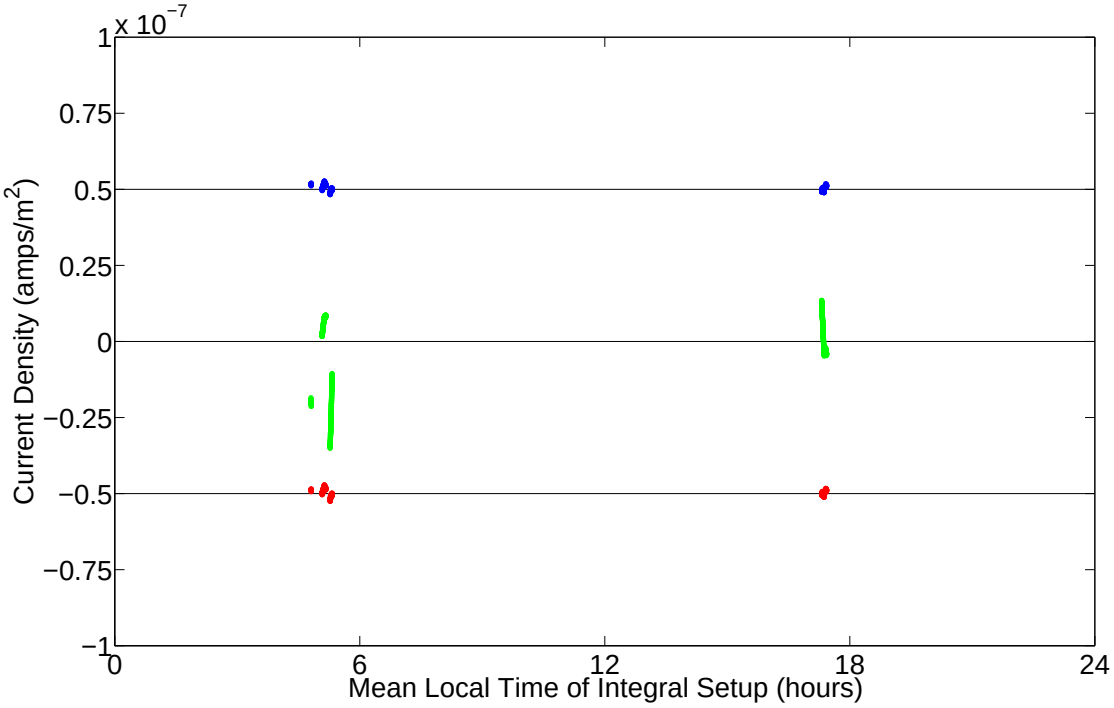

Figure 3.23: Simulated *Swarm* data: solutions for zonal current density. The distribution of *Swarm* overlaps between the upper and (one of the) lower satellites in LT versus current density is repeated three times, with arbitrary vertical offsets for clarity. The ordinate scale is identical to that of Figure 3.20 for ease of comparison. The blue data show the unedited solutions, exhibiting minimal current density. The red data show solutions from the same overlap series, with an error of 10 arc seconds applied to the first Euler angle of the lower satellite – the effect on the solution is minimal. The green data again show solutions from the same overlap series, with an error of 431 arc seconds applied to the first Euler angle of the lower satellite – the effect is appreciable. The three black lines are zero-lines of current density for each of the three series.

demonstrating an appreciable effect, are shown by the green points in Figure 3.23. Hence to mimic the magnitude of the current densities obtained with Ørsted and CHAMP, the application of unrealistic attitude errors is required. From this I infer that typically occurring attitude errors will have an insignificant effect on the results.

The Euler angles are estimated in-flight. Convention for the timescale of this estimation differs, but is typically on the scale of at least a day (and longer for Ørsted than CHAMP (Olsen et al., 2006b)). The Ørsted and CHAMP overlaps occur on each 'side' of the Earth at the same UT, and each epoch of overflights occurs over a period of roughly 15 days. Whilst I have ruled out the possibility of attitude error directly causing the magnitude of current densities

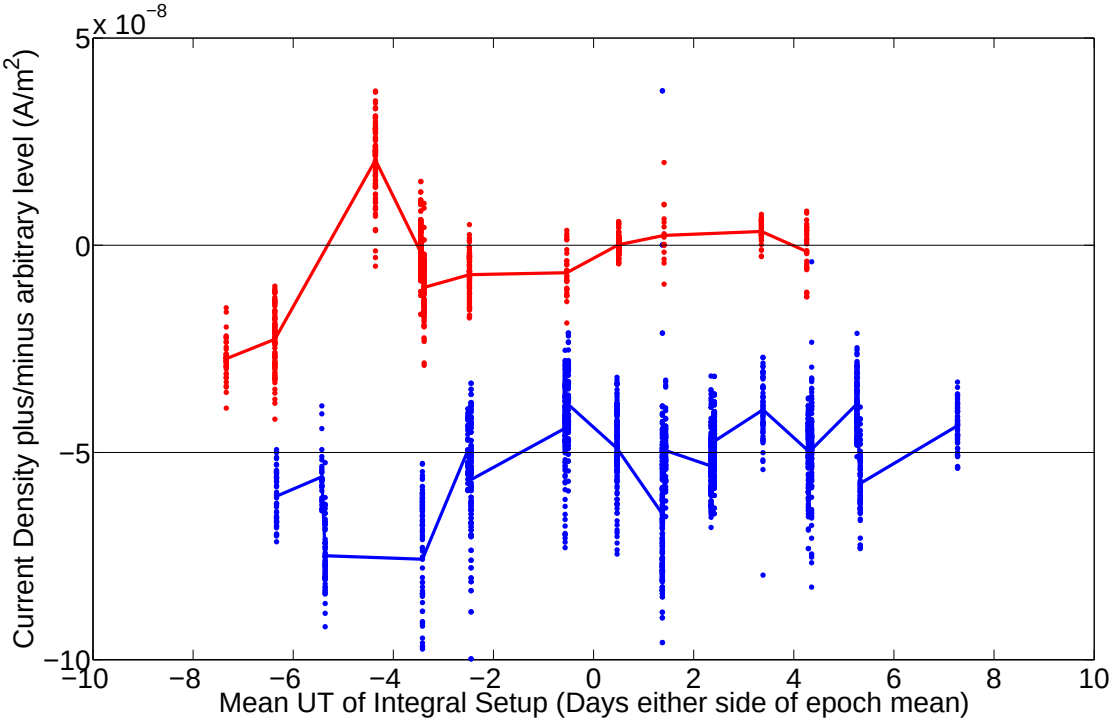

Figure 3.24: Comparison of current density estimates made at the same UT, shown in the red and dark blue coloured boxes in Figure 3.20 (the colors here are as per the colors of the boxes). The blue coloured values have been shifted in the negative direction by 0.05 μA/m^2 in order to distinguish the two data sets: the black 'zero' lines are at levels of 0 and -0.05 μA/m^2 for red and blue sets, respectively. A line has been traced through the mean of each cluster of values to highlight the trends in UT. A comparison of the shared UT trends in relation to the trends in Figure 3.26 (which shows radial line element length discrepancy) is made in the main text. In summary, since this set of overlaps shows no effect from Euler angle estimation, and a superficial similarity to the UT trends in the radial line element length discrepancy, the error from both these factors is likely minimal.

estimated, it does not preclude the possibility of the currents being aliased between opposing LTs (recorded at the same UT) by the Euler angle estimation process. Here I look at integral solutions which are similar in UT but separated in LT to assess the effect of the Euler angle estimation.

In Figure 3.24 I show the two sets of values in the red and dark blue boxes in Figure 3.20 (the colors here are as per the colors of the boxes). Here they are distributed in UT, with an arbitrary vertical offset of -0.05 μA/m^2 applied to the blue points to distinguish the two sets. A line has been traced through the mean of each cluster of values to show the trends in UT.

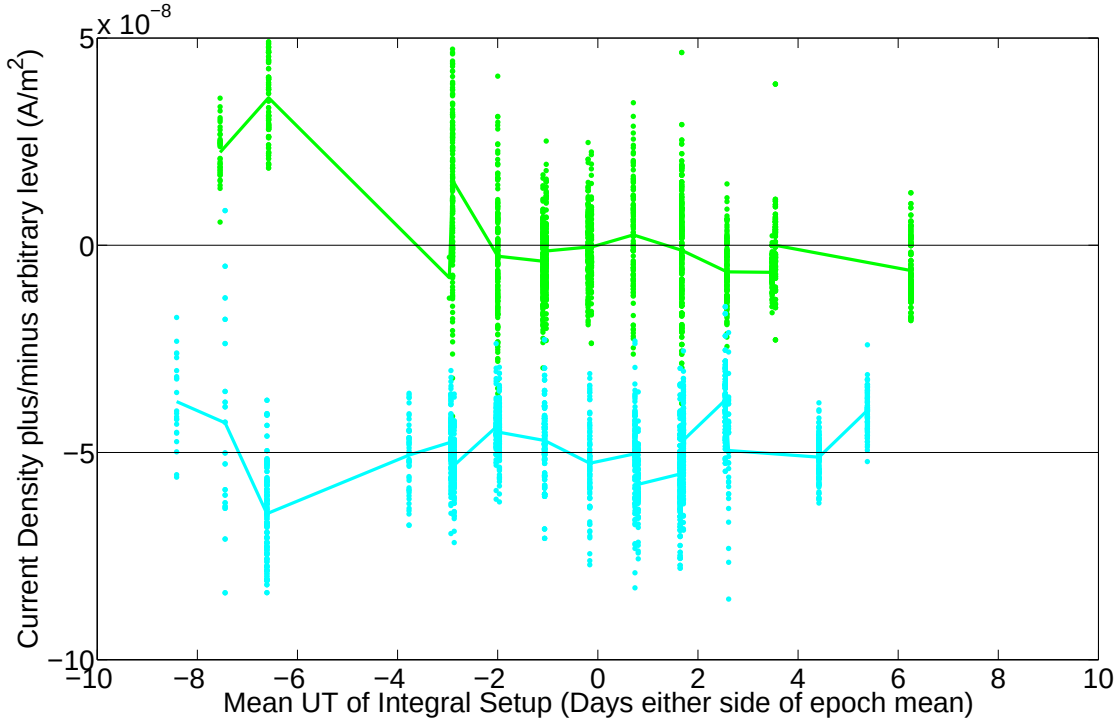

Figure 3.25: Comparison of current density estimates made at the same UT, shown in the green and light blue coloured boxes in Figure 3.20 (the colors here are as per the colors of the boxes). The blue coloured values have been shifted in the negative direction by 0.05 μA/m^2 in order to distinguish the two data sets: the black 'zero' lines are at levels of 0 and -0.05 μA/m^2 for green and blue sets, respectively. A line has been traced through the mean of each cluster of values to highlight the trends in UT. A comparison of the shared UT trends in relation to the trends in Figure 3.27 (which shows radial line element length discrepancy) is made in the main text. In summary, this set of overlaps shows a strong impact from the Euler angle estimation process but shows no similarity to the UT trends in the radial line element length discrepancy.

Both red and blue sets show a broad trend for values increasing with UT. Each cluster of values occurs within a few minutes of UT, so if the Euler angle estimation was affecting these values, a 'mirrored' trend (opposite in sign for each opposing LT) would be seen in their UT progression. No such trend is seen in this example, but this is not the case for all epochs. In Figure 3.25, I show the two sets of values in the green and light blue boxes in Figure 3.20, in a similar layout to Figure 3.24. This set of overflights does appear to be affected by the Euler angle estimation. Note that the combination of the CHAMP and Ørsted orbital inclinations (and other defining parameters) causes the overflights to precess through colatitude and longitude with UT. The progression of the trends in Figures 3.24 and 3.25 with UT is due simply to the precession through colatitude, and is not (principally) related to any actual change in the currents with UT.

In the set of overflights shown in Figure 3.25 it is not possible to separate which of the two LTs the current density trends stem from as the magnitudes are split evenly between both opposing LTs. Fortunately, most sets of overflights appear unaffected by the Euler angle rotation process, and in this chapter I have only analysed current density estimates from LT sectors which do not exhibit this aliasing effect.

In chapter 1 I discussed the aliasing of external field signal by the CHAMP data attitude-calibration process, resulting from the periodic in-flight re-estimation of the Euler angles. The timescale for the re-estimation differs, but is presumably tied to the degree of thermo-mechanical bending of the magnetometer's instrument bench. Hence, the error from this bending is quantified. It is possible that the LT-dependency (shown in Figures 3.25 and 3.24) of the aliasing introduced by the re-estimation could be factored into the timescale on which the Euler angles are estimated, for an overall more accurate approach to the orientation problem. Whilst at present it is not clear why certain LTs are affected by the aliasing more than others, a long timescale for the Euler angle estimation (*i.e.* meaning that CHAMP precesses through as many hours LT as possible) is desirable, as long as the error from the thermal bending does not exceed the error from the LT-aliasing.

3.6.3 Effect of satellite altitude changes within the integral calculation region

In the integral loop, the longest line elements are the radial lines connecting the Ørsted and CHAMP arc edges. I treat the magnetic difference across this gap in the same manner as that across any of the other line elements in the circuit. Any difference in the length of the two radial line elements could lead to a multiplication effect in the net magnetic difference, biasing the integral summation. Here I assess this effect using the same two sets of values as shown earlier in Figures 3.24 and 3.25.

The UT trends in the radial line element length discrepancy for the two sets of data in Figure 3.26 are nearly identical. Whilst the slope of the UT trend in Figure 3.26 appears superficially similar to the UT trend in the current density estimates shown in Figure 3.24, note

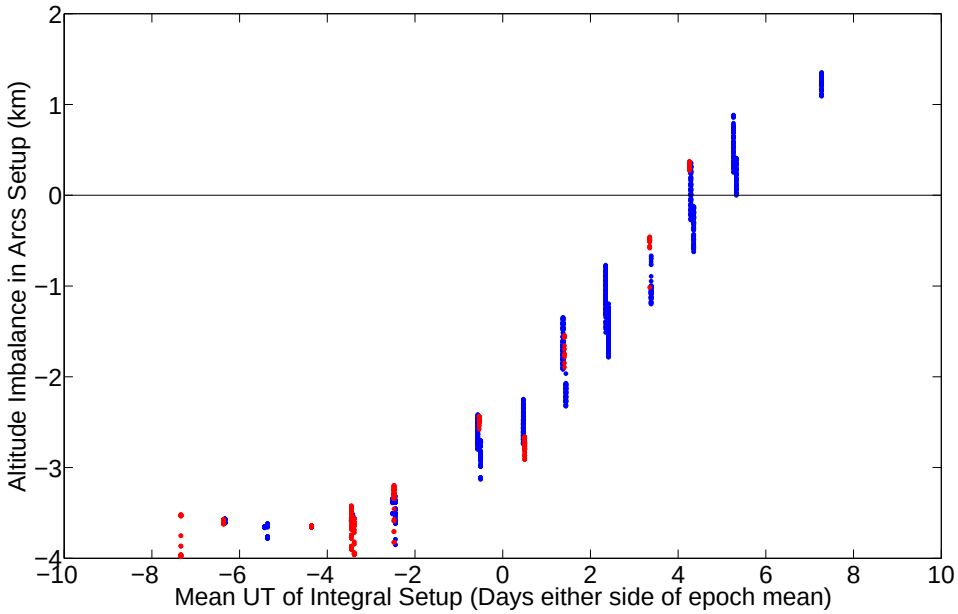

Figure 3.26: Radial line element length discrepancy in units of km, for overlaps in the red and dark blue boxes in Figure 3.20 (the colors here are as per the colors of the boxes). Each point is the difference between the radial line element lengths in each integral loop. The two trends in UT here are very similar, and bear a superficial similarity to the UT trends in Figure 3.24. Further analysis of the trends is given in the main text.

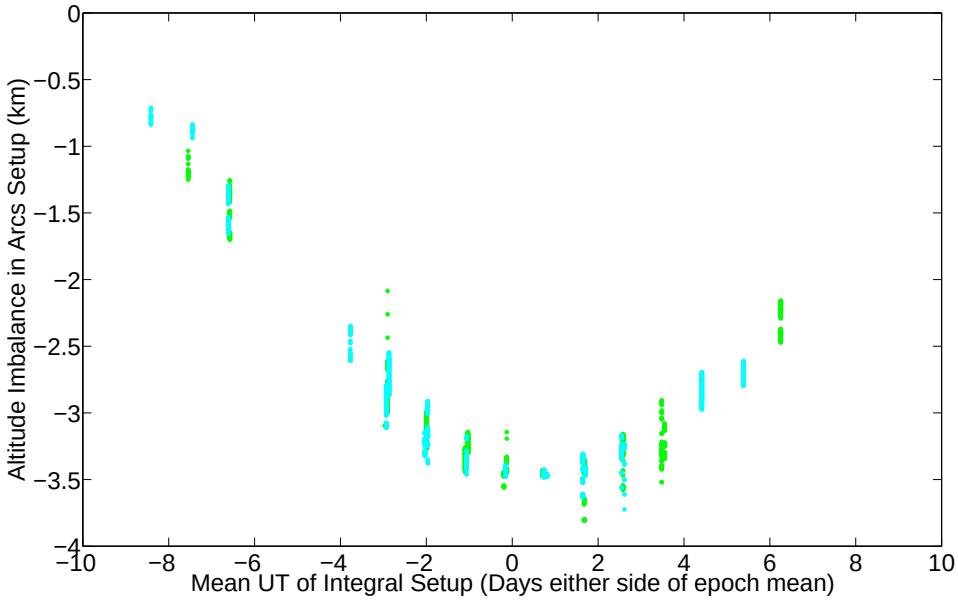

Figure 3.27: Radial line element length discrepancy in units of km, for overlaps in the green and light blue boxes in Figure 3.20 (the colors here are as per the colors of the boxes). Each point is the difference between the radial line element lengths in each integral loop. The two trends in UT here are very similar, but bear no similarity to the UT trends in Figure 3.25. Further analysis of the trends is given in the main text.

that the trends in the radial line element length discrepancy are nearly maximal at the UT of 0, whereas the same is not true for the current density estimates. The two UT trends in radial line element length discrepancy shown in Figure 3.27 are also very similar to each other. The same is not true of the current density estimates for the same data, shown in Figure 3.25. These results indicate that a discrepancy in the lengths of the radial line elements does not appear to affect the estimation of current density.

3.6.4 Assumption of vertical current trends across calculation region

In section 3.4, I made the assumption of current density varying linearly with altitude throughout the calculation region. The altitude distribution of the current density predicted by CTIP is shown in Figure 3.18. Rather than varying linearly between the two satellites' orbits, the fall-off of the current density with altitude is approximately exponential. My current density estimates are thus biased from the real-world case by an amount equivalent to the difference between the real gradient in the current density across the calculation region, and the assumed linear gradient. I have constructed a synthetic test of input and recovered model values of current density to illustrate the degree of bias this simplification imposes, which I summarise here.

The data used in the synthetic recovery test are the Ørsted and CHAMP measurement positions for the satellite data arcs in a series of overlap locations (those in the green box in Figure 3.20). There are over 3000 overlaps in this set of data, each of which (initially) comprises an arc of Ørsted data 2° colatitude in length, and the same for CHAMP. For each overlap, I construct two integral loops of location data. The first loop is simply the positions of the satellite data comprising the overlap, with no measurement locations on the radial line elements – this is termed a 'sparse' loop. For the second loop, the mean sampling interval of the Ørsted and CHAMP measurements is used to synthesise a series of 'measurement positions' along the radial line elements of the sparse integral loop. The combination of the sparse loop and these new data positions is termed the 'dense' loop, intended to represent an idealised data distribution for the integral.

I sampled 3D CTIP current density values to each of the data positions in the 3000+ sparse and dense loops. For each loop the mean current density in the direction normal to the plane enclosed by the integral loop was calculated – sparse loops have a slightly higher mean current density than dense loops due to the aforementioned non-linear fall off in current density with altitude. The mean current densities were used to calculate the synthetic current I_{syn} enclosed in each integral loop by multiplying the appropriate mean current density by the area enclosed in the integral loop. This area was calculated from a summation of the triangular area elements of the sparse loops, as given by equation (3.15) in section 3.4.3. The enclosed area is identical for the sparse and dense loops. The synthetic current (sparse/dense respectively) is then used to calculate its magnetic effect at each of the sparse and dense loop data locations via an application of the Biot-Savart law (Fleisch, 2008, page 47)

$$\mathbf{B}_{syn} = \frac{\mu_0 I_{syn}}{2\pi r}\hat{\mathbf{s}}, \tag{3.18}$$

where $\mathbf{B}_{syn}$ is the magnetic effect (at a certain location) of the enclosed synthetic current I_{syn}, μ_0 is the permeability of free space ($4\pi.10^{-7}$ T m/A), $\hat{\mathbf{s}}$ is a unit vector in the direction of the magnetic field which results from a current flowing through an infinite-length thin wire at the centre of the integral loop, aligned in the direction normal to the plane of the enclosed area, and r is the length of a vector connecting the centre of the integral loop to the point at which $\mathbf{B}_{syn}$ is predicted. This application of the Biot Savart law (as well as the previous step of taking the mean current density for the integral loop) reduces the original 3D current density distribution to a mean scalar value. This is the same simplification implicit in my application of the integration method used to resolve current density from the real data, so the errors resolved in this test should be similar to those encountered in the real data. $\mathbf{B}_{syn}$ is the synthetic data equivalent of $\Delta\mathbf{B}$ in equation (3.4), and thus the line elements, along-track rotation, enclosed area and resulting recovered synthetic current density are calculated via the process described for the real data in section 3.4.

The input and recovered current density values for the sparse and dense loops are shown

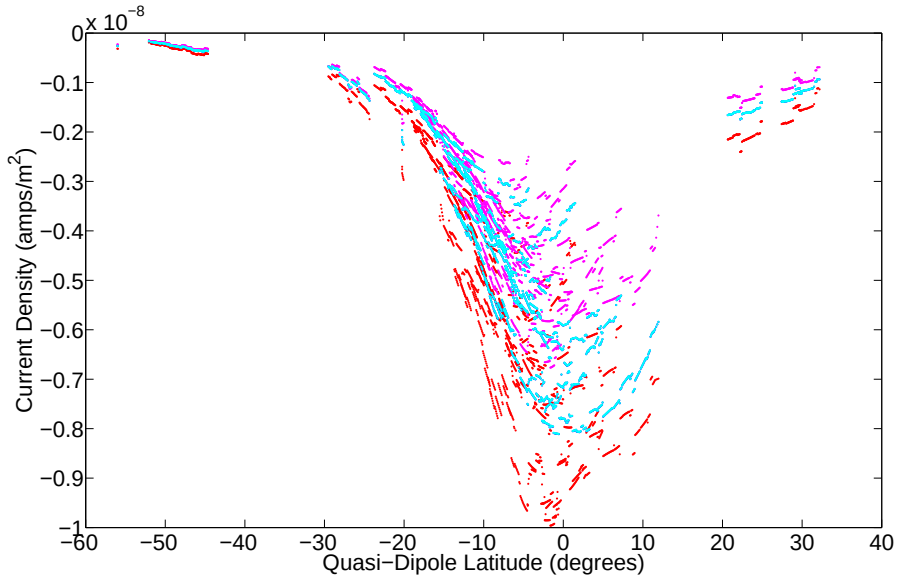

Figure 3.28: Values of input and recovered synthetic current density model prediction data for a series of overlaps at the same LT. The input values of current density prediction means for the sparse and dense integral loops (explained in the main text) are shown as red and dark blue points respectively. Note that the dark blue points are obscured. The magenta and light blue points are the recovered current density mean values for the sparse and dense loops respectively. The proportional recovery is shown in Figure 3.29.

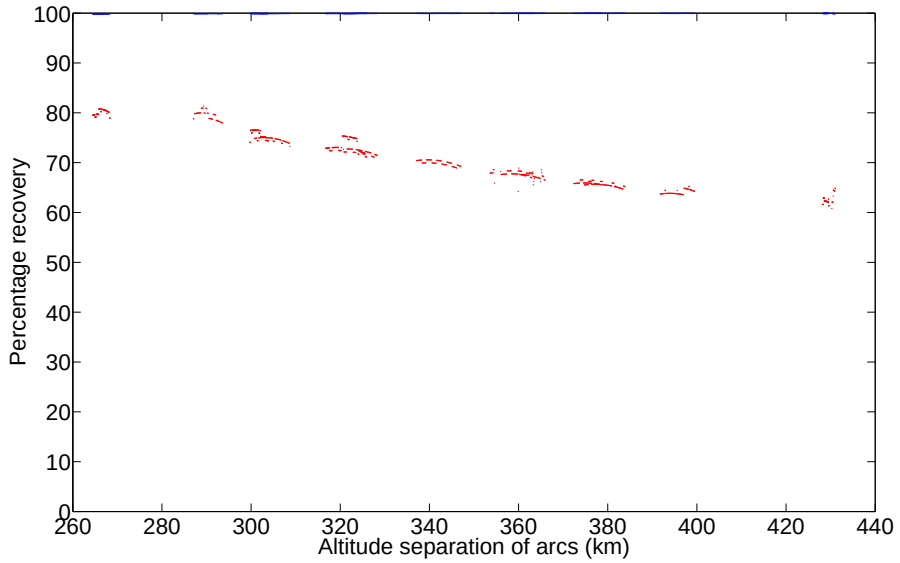

Figure 3.29: Percentage recovery of synthetic current density model prediction data for a series of overlaps at the same LT, shown plotted against the altitude separation of the two satellites. Red points are the percentage recovery for sparse integral loops, blue for dense loops (the terms 'sparse' and 'dense' are explained in the main text). The dense loop recovery values do not drop below 99.98%. The sparse loop recovery varies between 80 and 60%, dependent upon the altitude separation of the Ørsted and CHAMP satellites. The range of altitude spans shown is representative of the full extent of altitude spans seen in the real data.

in Figure 3.28. The input values for the dense loops are obscured by the dense loop recovery values, since their retrieval proportion is near 100%. Figure 3.29 shows the proportional recovery in more detail. The sparse loop recovery proportion varies between 80 and 60%, dependent upon the altitude separation of the Ørsted and CHAMP satellites. The sparse loop recovery is considered representative of the case for the real data, and the estimates presented in section 3.5.1 are assumed to be slight under-estimates of the real case. However, as stated in section 3.4.4 this does not affect my interpretation of the current density trends, which are considered robust. In addition to the discussion of vertical current density trends here, I note that the horizontal variation in the current density (shown in Figure 3.18) across the calculation region (2° colatitude) is negligible, and the along-orbit propagation of the integral region should offset any errors resulting from horizontal gradients in the current distribution. The assumptions in my model do not invalidate the results, as long as these limitations are borne in mind when drawing conclusions from them.

An improved description of radial ionospheric electrodynamics requires a more complete data set. ESA's upcoming mission *Swarm* (described in Olsen et al. (2007)), with its upper satellite at an altitude of 530 km and its two lower satellites at an altitude of (initially) 450 km, has an ideal vertical distribution for an independent resolution of these currents. However, after the initial launch of the satellites in the same orbital plane, the upper and lower satellites do not occupy similar LTs for another 3.5 years. This may be after the atmospheric re-entry of the lower pair of satellites. If the Swarm satellites do achieve more than a single series of crossovers, they will provide a more detailed altitude profile of the currents than the Ørsted/CHAMP configuration, but at a limited series of LTs.

3.7 Conclusions

I have demonstrated a robust method of resolving zonal current density on an event-by-event basis at LEO altitudes using satellite magnetic data. The use of multiple satellites minimises the contribution of magnetic fields outside the calculation region to the resolved current density, since no assumptions about current geometry or stationarity are required. Ampère's integral

is able to provide useful information about current flow at LEO satellite altitude, within the framework of the biases already stated. The data provided by this application of Ampère's integral span all LTs twice, but this data set cannot be used by itself to provide a consistent global representation of the currents. At any one overlap, the factors of satellite altitude, season, solar flux, missing data, and distribution in colatitude could combine to disrupt any trend that might otherwise be coherent in LT. Despite this, the results largely mirror the EIA cycle in LT, and appear remarkably resilient to the other factors affecting each overlap. The currents also appear unmodulated at times of either high-negative Dst or high $F_{10.7}$. This has implications for any future efforts to model the effect of these currents, and indicates that their generation is primarily associated with a steady diurnal pattern of atmospheric plasma generation and transport. Current densities in the range $\pm 0.1\ \mu A/m^2$ are resolved. The highest magnitudes of current density are typically seen in the equatorial regions, the result of Ørsted and CHAMP passing through the EIA in colatitude as the two satellites cross in LT – this is clearest between dusk and midnight, as the noise in the dayside estimates increases with LT. After midnight, once the peak density of the EIA has descended appreciably in altitude on nightside, I resolve significantly lower equatorial current densities with greatly reduced scatter. The mid-latitudes of the nightside LT sector are typically considered free of electric currents, yet here I consistently resolve a series of zonal current density intensifications. The cause of these unexpected intensifications remains an open issue. However, I suggest caution when using only nightside data in geomagnetic field modelling, as my results indicate that it may not be as free of electric currents as is usually assumed.

A comparison of the satellite measurements with CTIP current density predictions shows reasonable spatio-temporal agreement on the dayside, with both data sets consistently exhibiting strong westwards current flow at LEO altitude. The agreement on nightside is poorer at high latitudes, but both data sets show a significant magnitude reduction relative to dayside near the equator. CTIP has shed light on the balance of forces contributing to the total zonal current density, confirming that the EIA signal remains strong several hours into the nightside as shown by Maus & Lühr (2006). The magnitude of the satellite estimates are high relative

to the CTIP prediction. This is likely due in part to CTIP's prediction being for the 'quiet' ionosphere – I note that the agreement of my results with existing estimates of current density (*e.g.* Olsen, 1997) is within expectations. However, the independence of my current density estimates on geomagnetic field, and solar, activity (as measured by commonly-used indices) would seem to indicate that CTIP's predictions are too low in magnitude. Due to the apparent invariance of the current density estimates at changing solar and external geomagnetic conditions, identification of the ionospheric physics behind the difference in magnitude between my results and the CTIP prediction is beyond the scope of this study. However, the interested reader is referred to Alken et al. (2011) for a discussion on a possible cause of the pressure-gradient current's high strength, and to Kelley (2009), and Schunk & Nagy (2009) for a discussion on the competing electrodynamic influences in the F-region.

Chapter 4

Empirical Orthogonal Functions

4.1 Glossary

Table 4.1: Glossary of terms used in this chapter

Abbreviation	Explanation
EOFs	Empirical Orthogonal Functions
$\mathbf{X}$	Data matrix (columns are time series of single-component magnetic data)
$\widetilde{\mathbf{X}}$	Centred data matrix (columnar means removed)
$\widetilde{\mathbf{X}}'$	Centred data matrix, weighted for observatory distribution
S-mode	Method of arranging $\mathbf{X}$ as columnar time-series for spatial analysis
T-mode	Method of arranging $\mathbf{X}$ as row-vector time-series for temporal analysis
$\mathbf{R}$	Covariance matrix (S-mode, constructed from $\mathbf{X}^{\mathsf{T}}\mathbf{X}$)
$\mathbf{R}'$	Weighted S-mode covariance matrix, made from $\left(\widetilde{\mathbf{X}}'\right)^{\mathsf{T}}\widetilde{\mathbf{X}}'$
$\mathbf{V}$	EOFs; eigenvectors of $\mathbf{R}$ (the S-mode covariance matrix)
$\mathbf{v}$	Single column-vector of $\mathbf{V}$
$\mathbf{W}$	Spatial weighting matrix
$\mathbf{w}$	Spatial weighting values, a vector equal to diag($\mathbf{W}$)
$\mathbf{B}'$	EOFs of data weighted for station distribution; eigenvectors of $\mathbf{RW}$
$\mathbf{B}$	De-weighted EOFs of expansion of weighted data
$\mathbf{b}$	Single column-vector of $\mathbf{B}$
B	Single scalar element of $\mathbf{B}$
$\mathbf{U}$	Eigenvectors of the T-mode covariance matrix (constructed from $\mathbf{X}\mathbf{X}^{\mathsf{T}}$)
$\mathbf{T}$	PCs (Principal Components); temporal evolution of the spatially-static EOF patterns
$\mathbf{t}$	Single column-vector of $\mathbf{T}$
$\mathbf{Y}'$	PCs of data weighted for station distribution
$\mathbf{Y}$	PCs of data weighted for station distribution, scaled to units and range of $\mathbf{X}$
$\mathbf{y}$	Single column-vector of $\mathbf{Y}$
$\mathbf{L}$	Eigenvalues of $\mathbf{R}$ (the S-mode covariance matrix)
R_E	Earth's radius
A	Spherical triangle surface area
E	Spherical excess (residual of: sum of the spherical triangle's angles, and π radians).
D	Degrees of freedom of the data matrix.
mode	The pair of an EOF and its eigenvalue, forming a part of $\mathbf{X}$
LT	Local time (not latitude-specific; does not account for seasonal variation)
GEO	Geographic coordinate system (Hapgood, 1992)
MAG	Tilted-dipole geomagnetic coordinate system (Hapgood, 1992)

Acknowledgements

The work in the Empirical Orthogonal Functions (EOF) chapter was begun during a month-long research placement at DTU Space in Copenhagen, Denmark, during which time I worked with Nils Olsen on the start of the EOF study. The idea to use EOFs to resolve the external field signals important in mantle induction studies was motivated by an unfunded NERC grant submission, in which Kathy Whaler was a participant and Nils Olsen a named Project Partner. The work in this chapter was done by me, with advice on associated theory and practice provided by my supervisors, and by Nils Olsen. Matlab code for coordinate transformations and the spherical harmonic analysis was provided by Nils Olsen and modified by me. Matlab code used to produce Figure 4.3 was provided by Nils Olsen. My supervisors also helped with proofreading.

4.2 Motivation and objectives

One of the main findings of the Virtual Observatory (VO) study (chapter 2) was that whilst the *Swarm* constellation geometry does fractionally improve (in relation to a single-satellite case) the description of the external fields in a geographic reference frame, the improvement was much less than expected. From Beggan (2009, sections 4.4 and 4.5 for context, section 4.6.1 for the CM4 corrections) we know that the majority of the contamination affecting VOs is related to the symmetric ring current (and spatial aliasing thereof). Significant magnetic field contributions from Earth-external sources which are asymmetric in local time (LT) (*e.g.* from the partial ring current and ionosphere) were also demonstrated to affect the VO solutions – my thesis has confirmed this effect in chapter 2 with *Swarm* synthetic data from the E2Eplus model (Olsen et al., 2007).

The limited benefit from the constellation geometry indicates that in the GEO-fixed frame the mid-latitude external fields have a significant magnitude variation on time scales of less than six hours. That is the time (in UT) taken for two LT-separated *Swarm* satellites to sample the same geographic location (in this case, the same VO bin) when the satellites are at their

maximum separation in LT. Most of the variance during this time will be due to the change of the symmetric ring current, but not all of it. Due to the somewhat simplified parameterisation of the LT-asymmetric magnetic fields in the E2Eplus model (both in terms of UT and LT), I was not able to reliably determine how much of the total aliased signal was due to the LT-asymmetric fields. If we are to take full advantage of the increases in resolution and accuracy of satellite-sampled magnetic data when resolving GEO-fixed structures, then corrections for the external fields which take their UT and LT variation properly into account are required. Therefore, we need to know more precisely how much each geomagnetic field source contributes to unwanted signal in satellite data. This requires non-synthetic data and a constant time series of rapid sampling of all local times simultaneously (*i.e.* of period $\ll$ the 4.5 months offered by CHAMP). This cannot be achieved with currently available satellite data, and so in this study I use hourly means data from ground-based permanent magnetic observatories.

This study aims to describe the distribution of the external fields in LT in more detail. The focus is on how the LT-fixed external magnetic field at mid to low latitudes changes spatially and in terms of magnitude across a solar cycle, in the tilted-dipole (MAG) coordinate system (viewed in the LT frame). Note the use of the term 'spatial' to describe the form of the daily variation signal in the LT frame, and that the term 'daily variation' here includes more than ionospheric signal. Figure 1.4 (in chapter 1) shows a schematic diagram of the magnetosphere. Note the local time distribution of the magnetospheric current systems. The symmetric ring current, the magnetic field of which is the q_1^0 term of equation (1.1), affects all local times equally in MAG. I thus say that it is zonally continuous in the LT frame, or 'LT-symmetric'. I describe the magnetic fields of current sources which are not zonally continuous in the LT frame as being 'LT-asymmetric' – at mid-to-low latitudes this will include the magnetopause, magnetotail, neutral sheet and partial ring currents, as well as the ionospheric current systems. The magnetopause current peaks at local noon, whilst the magnetotail and plasma sheet (neutral sheet) currents have their peak effect at local midnight (Russell, 2000). An overview of the partial ring current is given by Campbell (2003, page 144) and Kivelson & Russell (1995, pages 289 and 409). It is assumed to close in the ionosphere at high latitudes via field aligned

currents connected to the ends of the partial ring, though in-situ current observations made with measurements from the Cluster mission of magnetospheric satellites infer other causes – this is still an open issue (pers. comm. H. Lühr). Le et al. (2004) showed that the ends of the partial ring (*i.e.* the FACs) are not at noon and midnight but rather at dawn and dusk. This causes the peak effect of the partial ring current to be between dusk and midnight, moving duskwards with higher field activity. Le et al. (2011) discussed the dominance of the partial ring current over the symmetric ring current during magnetic storms, though I only consider quiet times in this study. Le et al. (2004) also showed the symmetric ring current to have a minimal signature on dayside, which I will later confirm (section 4.8.5). The use of ground-based observatories means that ionospheric currents will affect the distribution of magnetic signal in LT. From the ionosphere a strong noon intensification is expected, and possible dusk intensifications if the currents resolved in the Ampère's integral study (chapter 3) have measurable impact at ground level. The rotation of the Earth underneath these current systems produces predictable patterns in magnetograms recorded at the Earth's surface.

I focus on the LT-asymmetric external magnetic fields for the following reasons. Studies of these fields with the VO method (Mandea & Olsen, 2006; Beggan, 2009) have shown that the LT-asymmetric fields are a concern in core field modelling from satellite data, having the potential to affect models such as gufm-sat (Finlay et al., 2012). The improved description of the long-period external field that I will produce in this study will be useful in the effective utilisation of the *Swarm* data. Mantle induction studies often rely on a simple zonal geometry assumption (*i.e.* just the q_1^0 symmetric ring current, as discussed by, *e.g.* Olsen, 1999; Kuvshinov & Semenov, 2012) for the external-field inducing source description. Mantle induction is a key part of the *Swarm* science mission (Friis-Christensen et al., 2006), and better knowledge of the dominance of the q_1^0 symmetric (ring current) term in relation to the LT-asymmetric fields should aid these studies.

My results will have special relevance to mantle induction studies, so I focus mainly on long-period signals in my analysis (I define how long these periods are in later sections of

this chapter). The focus on these fields throughout a solar cycle is motivated in part by analysis (Campbell, 2003, page 143) which suggests that the solar cycle effects on the magnetic field are non-linear, differing from a simple increase in field activity – that the prolonged intensification of the external field sources changes them in a manner different to that occurring during a (relatively) short-duration geomagnetic storm. There would be effects from stochastic, enhanced field activity events, but my use of data selection techniques for quiet days should minimise the impact of these. It is also noted that *Swarm* is planned for launch in solar maximum.

When applying a spherical harmonic analysis (SHA) to geomagnetic data we obtain a differentiable function representing the potential field at all locations for the time-instant of the input data, based on a series of waveforms which are harmonic on a reference sphere (*e.g.* Campbell, 2003). It is common to represent the time-changes in an external field source-region or source-process by the (often regularised and spline-fitted) progression in time of a spherical harmonic, or group of harmonics (*e.g.* Sabaka et al., 2004). This approach carries the strong implicit assumption that the same set of harmonics provides an equally-good representation of the magnetic field of the source region of interest throughout the timespan of the analysis. The disadvantage of the use of a SHA for an analysis of this type will be demonstrated later in a case study using real-data – for now I focus on a pertinent existing example, in the form of the CM4 model (Sabaka et al., 2004).

Magnetic field models like CM4 have shown good data fits by UT-modulation of a relatively simple external field source spatial structure via time-series indices of solar and global geomagnetic activity. The simplicity of the external field source-region structures in LT is not disputed – these sources rarely exceed SH degree 3 in complexity at the Earth's surface – but the assumption that these patterns are spatially identical in all field conditions is certainly a concern. To decompose the external magnetic field on the scale of a full 11-year solar cycle, I avoid spherical harmonic analysis (SHA) and use Empirical Orthogonal Functions (EOFs). EOFs are a data-based signal processing method of analysing time series in terms of spatial patterns and periodicities. They 'project' the data onto patterns (directions) of maximum vari-

ance in a data set. The method overcomes some of the problems associated with assuming the nature of the source fields (*e.g.* that they have certain periodicities, or certain geometries) since the decomposition of the data arises directly from the information contained in it, rather than from *a priori* assumptions.

The main research questions of this chapter are as follows:

- Can EOFs provide a useful decomposition of the external geomagnetic field?
- Can the EOF decomposition be used throughout a full solar cycle to assess the change in the magnitude and LT-sector dominance of the LT-asymmetric external fields?
- Does this improved description warrant a change in the standard assumption of the inducing field geometry used in mantle induction studies and hence how these studies are approached?

4.3 Generalised definition of EOFs and S-mode EOFs

EOF analyses are commonly used in climatology as a method of statistically representing spatially and temporally coherent signals in a large, highly variable data set. The aim of the EOF analysis is to allow the separation of the data into a signal-subspace, and a noise-subspace (von Storch & Navarra, 1999; von Storch & Zwiers, 2002, pages 231, 294 respectively). The signal-subspace represents the dynamical behaviour of the patterns I am attempting to represent (Bjornsson & Venegas, 1997, page 13). The term 'noise'-subspace is slightly misleading in this case. The definition of the noise-subspace has nothing to do with the error or uncertainty on the input data, but is simply the part of the input data which does not significantly or usefully describe the process I am attempting to isolate – it contains the variance of the rest of the input data set. It is possible for the noise-subspace to contain some dynamical signal from the same physical process which contributes to the signal-subspace, but this typically arises when the EOF analysis has been constructed poorly with regards to the signal we are attempting to isolate. I describe what signals the EOF analysis is most effective at isolating in section 4.4 and demonstrate the EOF analysis set-up in section 4.6. Both signal- and noise-subspaces are

affected by the uncertainty on the measurements, as well as random perturbations in the data set (von Storch & Navarra, 1999, page 231 (start of section 13.1)). A generalised derivation of EOFs (and the specific variant which I use) is described below, based mainly on Bjornsson & Venegas (1997, page 12), von Storch & Zwiers (2002, pages 294-295) and Jolliffe (2002, page 5).

Consider a two-dimensional data matrix **X** (described in equation 4.1) populated with scalar data, the covariance structure of which is of interest to us, and which I will analyse with the EOF method. The arrangement of the elements of **X** has a strong effect on the output of the EOF analysis. Richman (1986, page 294) described six possible arrangements of the data matrix **X** for the generalised case of a set of time series representing some underlying scalar field, recorded at a distribution of stations. Of those 6, only 'S-mode' (Spatial-mode, which averages temporally to highlight spatial structure) will be used here, and will be discussed briefly in relation to 'T-mode' (Temporal-mode, which highlights temporal structure (Beckers & Rixen, 2003, page 1840)). I initially used S-mode type EOFs because, as discussed by Richman (1986), they respond in an appropriate manner (for my purposes) to the post-EOF processing 'rotation' method (refer to section 4.9.3). S-mode also averages temporally rather than spatially, and was better suited than T-mode to resolving the spatial patterns of long-period signals in the magnetic data. In S-mode EOFs the data matrix **X** is of dimension n x p, where p is number of stations, n is number of time-records. EOFs are composed of a series of p eigenvectors defining patterns which, when projected onto the data from which the EOFs are defined, successively maximise the variance of those projections – more detail is given later in this section. T-mode analysis simply transposes the arrangement of **X**. In this study the columns of S-mode **X** are the time series of n hourly means of a single component recorded at each of the p observatories:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \tag{4.1}$$

I am interested in partitioning the variance of $\mathbf{X}$, so prior to the EOF analysis of these data I remove the time mean of each column (data corrections applied prior to this step are given in section 4.6):

$$\widetilde{\mathbf{X}} = \mathbf{X} - \overline{\mathbf{X}} \tag{4.2}$$

where $\overline{\mathbf{X}}$ are the columnar means of $\mathbf{X}$, averaging over time. This process is called 'centring' the data matrix. As discussed by Jolliffe (2002, page 389), if the data matrix is not column-centred then the EOFs do not 'project' onto the variance, instead through some other origin, and the interpretation of the physical meaning of the EOF patterns is then less straightforward.

The covariance matrix $\mathbf{R}$ (dimension p x p) is formed from $\widetilde{\mathbf{X}}^{\mathsf{T}}\widetilde{\mathbf{X}}$, and the EOFs $\mathbf{V}$ are the eigenvectors of $\mathbf{R}$. Each EOF $\mathbf{v}$ (a column of $\mathbf{V}$) is a different linear recombination of the elements of $\mathbf{X}$, describing the structure of its covariances. The first EOF is that static pattern which 'explains' most of the variance of the input data. Each EOF is defined such that the pattern it represents is orthogonal to all others, and that the projections of these eigenvectors onto the original data are mutually uncorrelated. To find the first EOF $\mathbf{v}_1$, I maximise (subject to the criteria that $||\mathbf{v}_1|| = 1$, and $\mathbf{V}^{\mathsf{T}}\mathbf{V} = \mathbf{I}$)

$$\text{var}\left(\mathbf{v}_1^{\mathsf{T}}\mathbf{X}\right) = \mathbf{v}_1^{\mathsf{T}}\mathbf{R}\mathbf{v}_1 \tag{4.3}$$

where $\mathbf{I}$ is the p x p identity matrix and $\left(\mathbf{v}_1^{\mathsf{T}}\mathbf{X}\right)$ defines the projection of the first EOF onto the data matrix. Equation (4.3) also yields $\mathbf{v}_1^{\mathsf{T}}\mathbf{R}\mathbf{v}_1 = l_1$, where l_1 is the largest eigenvalue of $\mathbf{R}$, and $\mathbf{v}_1$ is the corresponding eigenvector of $\mathbf{R}$. In this study I refer to the eigenvectors of the covariance matrix as the EOFs, and their projections onto the original data $\left(\mathbf{V}^{\mathsf{T}}\mathbf{X}\right)$ as the PCs (Principal Components). There is one PC series for each EOF-eigenvector, and the pairs of these are termed the EOF 'modes' of the decomposition.

In S-mode EOF analysis the set of EOF basis vectors are each p-dimensional, representable as a spatial pattern which is static (a standing oscillation) over the timespan of the input data. The PCs (each n-dimensional) are then a set of associated time series describing the temporal evolution of these static patterns – their oscillation through the timespan of the input data. In S-mode analysis the PCs can be thought of as a time series showing the relevance of the EOF pattern to the input data's structure at any instant.

R is the 'temporal' covariance matrix, so-called because the resulting EOFs will project onto temporally-averaged spatial patterns. I find the eigenvectors and eigenvalues of **R** by solving the eigenvalue problem (given by Bjornsson & Venegas (1997)):

$$\mathbf{RV} = \mathbf{VL} \tag{4.4}$$

where **R** is the covariance matrix (dimension p x p), **V** contains the EOFs, or eigenvectors (dimension p x p) arranged column-wise, and **L** is the diagonal matrix of their eigenvalues (dimension p x p). Equation (4.4) is equivalent to equation (4.3) – I include it as it defines the EOFs construction more clearly and will be useful in later discussion (section 4.6.2). The eigenvalues represent proportional measures of the total variance in **R** explained by each EOF (Bjornsson & Venegas, 1997, page 9). For ease of reference, I define a method for computing the temporal evolution series as (Bjornsson & Venegas, 1997; Jolliffe, 2002, pages 15, 30–31 respectively)

$$\mathbf{T} = \widetilde{\mathbf{X}}\mathbf{V} \tag{4.5}$$

where the columns of **T** are the temporal evolution series (PCs). The data matrix $\widetilde{\mathbf{X}}$ is exactly the sum of the products of the EOFs and their temporal evolutions, leading to the 'reconstruction' equation (Bjornsson & Venegas, 1997; Jolliffe, 2002, pages 15, 30–31 respec-

tively)

$$\widetilde{\mathbf{X}} = \mathbf{T}\mathbf{V}^{\mathsf{T}} \tag{4.6}$$

Equation (4.6) can be written as (Bjornsson & Venegas, 1997)

$$\widetilde{\mathbf{X}} = \sum_{i=1}^{p} \mathbf{t}_i \mathbf{v}_i^{\mathsf{T}} \tag{4.7}$$

where $\mathbf{t}$ is a single PC (column vector of $\mathbf{T}$), $\mathbf{v}$ is a single EOF (column vector of $\mathbf{V}$), and p is the number of stations. The use of this alternative form is as follows. Having decomposed the input data into EOFs, eigenvalues and temporal oscillations, it is useful to reconstruct the parts of the original data describing the space-time variation of the patterns of interest (the signal-subspace referred to earlier). This is achieved by truncating the full summation of modes in equation (4.7) at some $i \ll p$, reflecting the signal-noise subspace transition point (Jolliffe, 2002, pages 30-31). The choice of i is clearly an important one. Many statistical measures have been developed to justify the truncation of the EOF description of the data – some will be shown later which focus on the control of the signal subspace patterns by the noise present in the data. However, the choice of which mode to truncate at is always subjective.

EOFs are not new in their application to magnetic data. Xu & Kamide (2004) give an overview of the uses of the method in recent years, including studies of internal fields (used in conjunction with SHA by Golovkov et al. (2007)), polar-region currents, sunspot number prediction and UT-frame daily variation decomposition. The method of Balasis & Egbert (2006) is closest to this study. The authors used mid-latitude nightside hourly mean magnetic data from 79 permanent observatories to study the nightside magnetic field variance in the LT frame, resolving a local time asymmetry focused over (local) dusk. I use a different data arrangement to Balasis & Egbert (2006) and my results are different to theirs, but compatible in the conclusions I draw from them. My study is new in that I assess longer time periods than have previously been studied using EOFs, and examine the effect of the solar cycle on the resolved

EOF patterns. Note that the use of magnetic vector data allows the use of three components in the analysis. EOFs do not hold any information about the spatial distribution of the input (the order of the p-axis of $\mathbf{X}$ is arbitrary, but consistent), and are not capable of treating the magnetic data vector as comprising component parts of the same underlying field. It is possible to apply the EOF analysis to the scalar magnitude of the magnetic field, but it is common in studies of mantle induction to use the horizontal components of the magnetic field to describe the inducing source, as this reduces the amount of induced signal in the resulting description. I run the EOF analysis separately for each of the magnetic field horizontal components. The validity of combining the output of these separate analyses will be discussed in a later section.

EOFs are a very general method, and can be used to decompose any system which can be usefully described by successively maximising variance on an orthogonal basis. A successive maximisation of variance will prove very useful in this study. I have no reason to suspect that the patterns I seek are organised on an orthogonal basis (*i.e.* are mutually physically exclusive), but will later show that this constraint does not impact the usefulness of my results. Justifying the interpretations I assign to the output of the EOF analysis I applied in this study cannot be discussed in full yet. It is useful though, to consider what structures the S-mode EOFs could project onto.

4.4 The meaning of S-mode variance

EOFs are ordered in terms of decreasing variance. von Storch & Zwiers (2002) termed the first mode "the most powerful single pattern" in representing the variance of the data matrix. The first EOF mode is simply the pattern which, when projected onto the data, has the most variance of any possible pattern produced from a linear combination of the elements of the data matrix, defined over the spatial domain of the input data (von Storch & Zwiers, 2002, page 295, section 13.1). The only reason the output is potentially useful is because the input is known to contain strong space-time structuring with a physical cause. This is reflected in the structure of the covariance matrix (which the EOFs define). Since the spatially distributed time series are not independent of one another, the covariance matrix has non-zero off-diagonal elements

(Jolliffe, 2002). However, that does not necessarily inform us as to what kind of spatial patterns should represent the strongest attractors in the EOF analysis. Here I assess which covariance structures are dominant in the representation of a field in the local time frame (which follows the sub-solar longitude at a geographically-consistent latitude) via an EOF decomposition.

Since the means have been removed from the input data (equation (4.2)), I now have a series of values which oscillate about their p separate zero-points. Recall that the use of S-mode EOFs averages the data temporally to pick out spatial patterns as eigenvectors, while the alternative T-mode analysis picks out structures in time. S-mode EOFs respond most strongly to spatial modes which represent standing oscillations, since these are collections of perturbations which repeat coherently spatially (Bjornsson & Venegas, 1997, page 6). Coherent variations tend to be dominated by low temporal frequencies and large spatial scale patterns, and such patterns will generally account for most of the variance of the system (Hannachi et al., 2007). It is this frequency-separation property which means the leading eigenvectors are described as the 'eigenspectrum' of the analysis. However, despite this useful property, EOFs are a poor choice for a frequency-based decomposition of a data set. If the spatial structure of a high-frequency signal correlates with the spatial pattern of the low-frequency signal in the leading mode, for example, then the temporal oscillation of this pattern will contain both high and low frequencies. The leading mode then describes a mixture of the physical sources of both different frequencies (this will later be demonstrated with real data (*e.g.* mode 1 in section 4.8.1)). Whether this mixture represents a distortion of the low frequency signal from random noise, or whether that spatial pattern contains a source which oscillates at a wide range of frequencies, depends on the variance associated with these sources. I cannot uniquely determine the physical inputs to each mode, though my interpretation can be aided by comparison of the mode's temporal evolution time series with an independent index (*e.g.* Dst, as demonstrated by Balasis & Egbert (2006)), and *a priori* knowledge of the characteristic frequencies of field sources with the same spatial pattern. Other indications of the effect of noise on the resolved modes will be demonstrated later.

The dominance of large spatial-scale patterns in the leading modes can also result from the orthogonality imposed in the EOF analysis. EOFs maximise variance over the domain of the input data. Since the orthogonality constraint is a global property, EOFs can exhibit large-scale structures as a result of more spatially-localised variance (Hannachi et al., 2007). In this case the EOFs are said to exhibit domain shape dependency (Richman, 1986). The domain dependency is most damaging when the first EOF is of a constant sign across the entire domain and of high amplitude, since this places a strong constraint on the possible shapes of the non-leading modes. Richman (1986) showed that, to an extent, the non-leading mode patterns for different covariance functions occur in the same predictable sequence. I discuss the impact of domain dependence on my results later (section 4.9.2) but note no strong effect. It should be clear (from the description of EOF projecting most strongly onto large oscillations) that the latitudinal oscillation of the seasonal signal in the LT frame represents one of the largest spatial signals I will resolve. It is common, in fact, to remove annual frequencies (cf. Bjornsson & Venegas, 1997; Balasis & Egbert, 2006) prior to the EOF analysis, so that it does not dominate the first mode. However, I have a focus on mantle induction (which relies on long-period signals to sound to the depths of interest) and thus apply no temporal filtering to my data set.

It remains to ascertain whether any useful aspects of the field are being overlooked by my focus on spatially-grouped variance - and if this affects any later physical interpretation of the resolved EOF coefficients. It is important to recognise that the EOF decomposition responds not to the absolute magnitude of the magnetic field, but to its tendency to vary in a spatially coherent manner. Therefore, signals which are static in the frame of reference throughout the timespan of the input data will be 'invisible' to the EOF decomposition, even if these represent strong signals in the (pre-centred) input data. EOF analysis is related to (the more commonly used in magnetic field analysis) SHA since both are projections of the data onto a set of orthogonally-defined basis functions (the harmonics in SHA, the eigenvectors in EOF analysis), thereby replacing the original data with a set of projection coefficients onto these basis functions. In EOF the projection coefficients are the PCs, in SHA, they are the Gauss coefficients (Eshel (2012, chapter 11, page 200), Preisendorfer & Mobley (1988)). However, each method

decomposes the data in a fundamentally different manner and cannot necessarily be used to represent the same information. A shortcoming of S-mode EOFs is that the temporal averaging of the spatial projections precludes the description of travelling waves by a single EOF pattern (travelling waves also affect T-mode analyses). The result of attempting to resolve a travelling wave is a pair of adjacent EOFs with near-identical eigenvalues, each describing some physically meaningless aspect of the same wave. Of the three types of coordinate-system-dependent field structures: stationary, propagating and oscillatory, EOFs can only resolve oscillatory patterns. As will be demonstrated in section 4.6, I construct the EOF analysis carefully to ensure that the patterns of interest oscillate in the frame of reference, and so, will be projected onto efficiently by the EOF decomposition.

4.5 The data set

I have relied on satellite data for much of the rest of this thesis. However, the EOF method requires that the measurement locations are spatially static throughout the span of the analysis. VOs would provide an ideal basis for spatially-fixing the satellite data in the GEO frame, but I have discussed at length the problems inherent in that approach in chapter 2, particularly when using a single satellite. Therefore, despite the reduced ability to determine if the external field signals are of ionospheric or magnetospheric origin, I adopt the use of ground based observatory hourly mean magnetic vector data in this analysis. The data are taken from a data set of 149 observatory time series, spanning the period 1997-2010, described in full in the ESA *Swarm* Level 2 processing report of Olsen et al. (2011). Figure 4.1 shows the distribution of the observatories in the GEO frame. Figure 4.2 shows the $F_{10.7}$ solar flux density index[1] over the span of the data set I use, showing the solar cycle I study.

The observatory data were prepared specifically for the purpose of resolving the high-degree magnetospheric field, and have been subject to a detailed quality control process, aimed at removing baseline discontinuities, data spikes and measurement errors in general. Though

[1] obtained from `ftp://ftp.ngdc.noaa.gov/STP/GEOMAGNETIC_DATA/INDICES/KP_AP/`

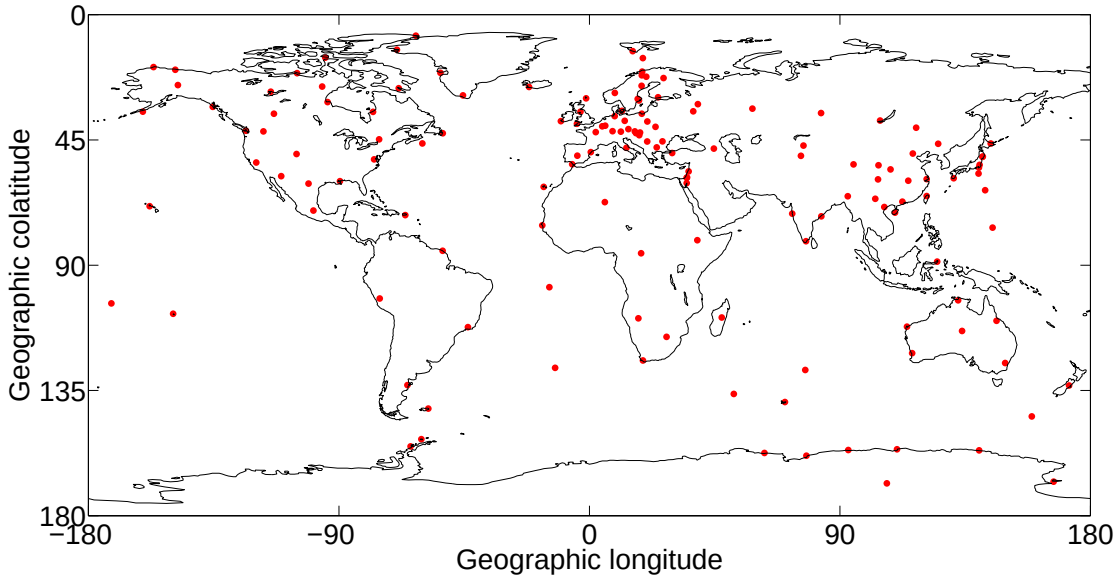

Figure 4.1: Geographic distribution of observatories from the study of Olsen et al. (2011).

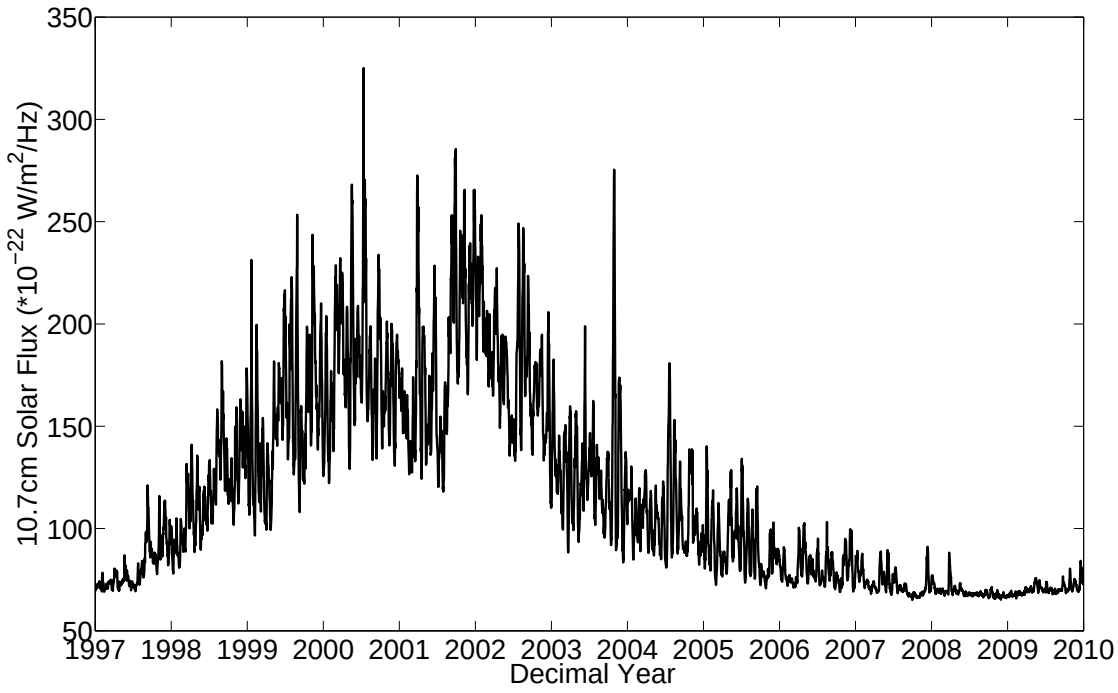

Figure 4.2: Solar flux density at 10.7 cm wavelength ($F_{10.7}$), as described in *e.g.* Covington (1969). Units are 10^{-22} W/m^2/Hz.

the study of Olsen et al. (2011) has a similar aim to mine, the authors used an hour-by-hour SHA to examine the magnetospheric fields, whilst I am using the EOF method as a more natural expansion of the data (the definition of the basis arises directly from the data).

In section 4.2 I briefly discussed the downsides of using an SHA in a study such as discussed in this chapter, including the temporal rigidity in the SHA basis. The fact that the spatial aspect of the field changes in time, whilst the spatial basis of the SHA does not, might seem a small sacrifice to retain the potential field description (which EOFs cannot). Here I examine that in more detail. Figure 4.3 shows a selection of the θ-component residuals (data minus model prediction) from a series of hour-by-hour global SHAs (each of degree 9, order 1) of the ground based observatory data's horizontal components. Omitting the radial magnetic field component still allows a full potential field description whilst reducing the effect of induced field signals, which I do not require in this example. Prior to each hourly SHA the data were corrected for core, SV, crustal, ionospheric and ionospheric-induced field estimates, leaving only the magnetospheric field contributions, which I resolve at a high-degree using the SH model. Figure 4.3 contains two main trends of note. Firstly, all residual traces exhibit an amplitude-modulation corresponding to the solar cycle, even at low MAG latitudes. Secondly, even in the mid-latitude region shown, the higher-latitude stations exhibit much stronger annual modulation of the residual signal. The likelihood is that much of the high-latitude residual signal is due to auroral electrojet and polar-cap currents. However, the presence of annual and solar-cycle based signal in the SHA residuals shows that the SHA is failing to retain information in two distinct (and useful) low-frequency bands, even at low latitudes. A more spatially complex SHA basis would incur less misfit, but would not be supportable by the observatory distribution. In summary, an application of a SHA prior to the EOF analysis (*e.g.* by Golovkov et al. (2007)) is useful for retaining the potential-field nature of the measurements, but in this case would probably have led to precisely the kind of spatial information loss I am using EOFs to avoid.

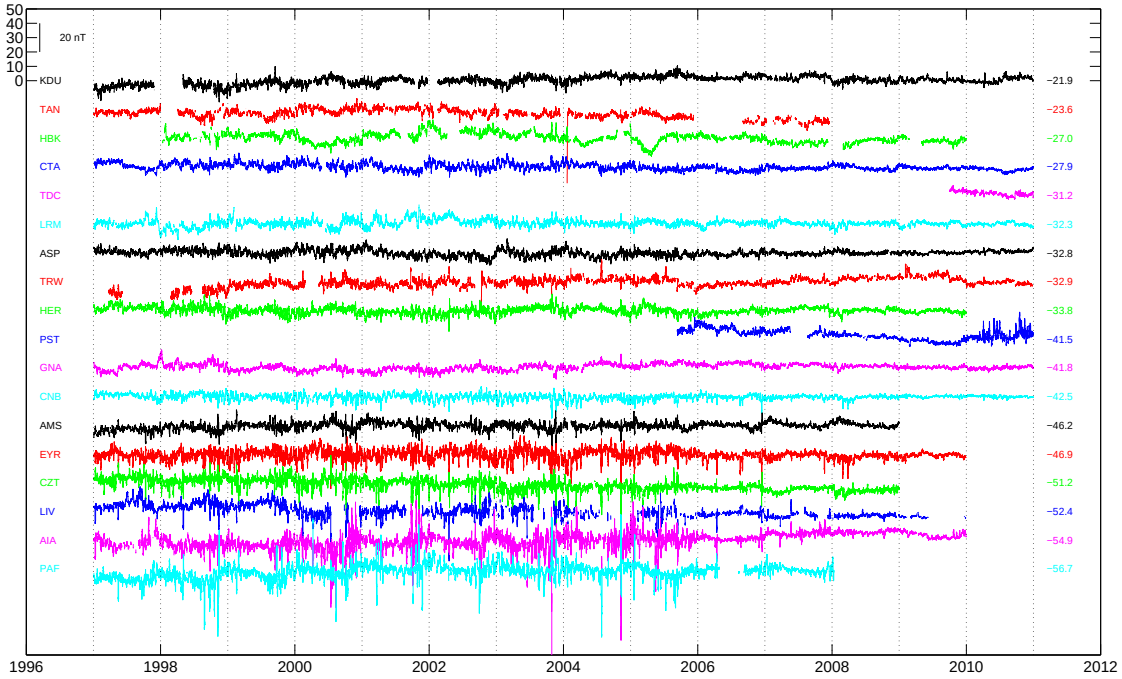

Figure 4.3: θ-component residuals from a global SHA of degree 9, order 1 of the observatory-sampled magnetospheric fields. The colours distinguish different observatory traces. For each station, the observatory acronyms are stated on the left, and MAG latitudes on the right. Note the annual modulation of the residuals at certain observatories – more detail is in the main text.

4.6 Setting up and applying the EOF analysis

The EOF analysis requires a number of preprocessing steps, listed below, which I will describe in turn in this section.

- Isolation of magnetic field contributions of interest
- Removal of data errors
- Subset data to certain temporal length
- Further temporal sub-setting – isolating one UT per day of data
- Infill missing values
- Select quiet days only
- Calculate a weighting metric for the station distribution

4.6.1 Pre-EOF processing: data selection, correction and interpolation

The first step in applying the EOF analysis is to isolate the field sources of interest to us. I am attempting to resolve departures from zonal symmetry in the MAG frame, so the data are corrected for core, SV, crustal and symmetric ring current magnetic fields. In a later section of this chapter, I assess the retention of the symmetric ring current signal in an EOF analysis, but it is removed in all analyses unless stated otherwise. A strong dayside signal is expected to result from the retention of ionospheric signal, but I will focus on the nightside in my analyses as I expect the partial ring current, which is strongest between dusk and midnight (Le et al., 2004), to account for much of the LT-asymmetric signal. The core field and SV estimates are removed using the CHAOS-4 model, the latest version of the CHAOS series described in Olsen et al. (2010c). The coefficients for this model are that of the preliminary model version CHAOS 4α (derived in December 2010)[2]. The magnetic vector data components and GEO coordinates are converted into the MAG coordinate system using geomagnetic pole locations for the 2005.0 epoch as given by the IGRF-11 magnetic field model (Finlay et al., 2010): colatitude 10.25°, longitude 288.19°. A coordinate transform using a variable pole position (dependent on the time between 1997 and 2010) was considered, but this would lead to stations which are not static in the frame of reference, and the different coordinate system would have no impact on the EOF output. Hence, epoch 2005 was used for the transformation as it represents the approximate midpoint of the 1997-2010 data span.

I then remove an estimate of the signal from the symmetric ring current using the vector data prediction of the RC index (both internal and external parts of the model, in the MAG frame) described by Olsen et al. (2011). RC data and model coefficient files were provided by Nils Olsen (pers. comm. 2011). The removal of the symmetric ring current signal is undertaken primarily to prevent longitudinally-symmetric signal dominating the first EOF mode. In section 4.8.5, I assess the effect of not correcting for the ring current in this manner – the results are informative, though not a main focus of this study, and the removal of the ring current signal is my default approach.

[2]obtained from `http://www.spacecenter.dk/files/magnetic-models/CHAOS-4/CHAOS-4alpha.mat` on 05/04/2011

The final step in the data correction is to remove the observatory bias which results from the short-wavelength lithospheric fields, which is not accounted for in the core field model. I follow the approach of Olsen et al. (2011) and calculate the mean observatory value for each component in geomagnetic quiet times, defined as when $Kp < 2^+$ and $|dDst/dt| < 2$ nT/hour. These mean values are removed from all values in the observatory time series. Earlier I noted the presence of discontinuities in the data. The quality control process applied by Olsen et al. (2011) involved splitting the time series of any observatory with a discontinuity which could not be corrected into two or more time series. I treated such instances carefully in my analysis, applying the baseline correction separately to each split-part of the observatory record, then merging them into one series for the observatory and visually checking the continuity of the merged series.

The corrected data set contains 149 observatories, but only 142 of these are deemed to have data of high enough quality for this study, due to unexplained baseline drifts which are unlikely to have a physical explanation (assessed via a set-up similar to Figure 4.3, not shown). For this reason I remove all data contributions from the observatories AAE, CMO, MIR, TUC, VOS, HUA and HBK. Parts of the time series of the observatories listed in Table 4.2 are also removed, but I retain most of the data from these stations.

Table 4.2: Observatory data removal notes

Observatory	Data problems
SBL	Small steps in the data during 2009
HLP	A small step/drift during 2006
VSK	Spikes in 2001
PPT	Noise during 1997-99, step/drift during 2000 and 2002
API	Drift during 2007
LRM	A spike during 2001
AIA	Spikes during 1997
AMS	Drift trend during 1997

The data set spans 14 years from 1997-2010 (though as discussed later in this section, I only use the first 13 years). If all the data are input into the EOF analysis at once, a signal of the change in the LT-fixed magnetic fields across the solar cycle will be obtained (one of the aims of this study). However, the use of S-mode EOFs means that this signal will 'smear' any spatial patterns which occur, for instance, only at the solar cycle peak. To avoid this issue I run the EOF analysis separately on 13 sets each containing 1 year of data, as well as on the data set of all 13 years.

Having derived sets of corrected hourly means data in MAG coordinates, I wish to look at the field variations in the magnetic local time (MLT) frame. MLT coordinates are time-variant, though MLT colatitude is identical to MAG colatitude. To define MLT longitude I must first define the SM (Solar Magnetic) coordinate system (Hapgood, 1992). The horizontal plane of SM is the MAG equatorial plane, and the SM y-axis is perpendicular to a plane containing the Earth-Sun line, and the MAG dipole axis. The vertical axis of both the SM and MAG systems are the dipole axis. MLT 'longitude' is measured in hours of local time, and is defined as (SM longitude-180)/15. I have discussed the problems inherent in using EOF analysis to resolve travelling waves. For similar reasons, I cannot apply the analysis in a coordinate frame in which the observatories are moving. Therefore, the data matrix $\mathbf{X}$ cannot be constructed from all UT hours of data – I must force the observatory distribution to be static in MLT throughout the timespan of the analysis, and I do this by selecting data at only one UT per day. The distribution of the observatories in MLT is then dependent on which UT is chosen to form my data subsets – this is termed the 'start-UT' of the data arrangement.

The choice of start-UT controls the clustering of stations at different local times, affecting the regions of best resolution in LT. The MLT station distributions at some start-UTs can be expected to provide a poorer description of the LT trends than that which would be obtained from another start-UT with a better station clustering. The start-UT also controls the placement of the GEO equator's latitude-tilt in MLT. Figure 4.4 demonstrates this effect. Note that changing the start-UT affects the proportion of each GEO hemisphere that is in each MAG hemisphere,

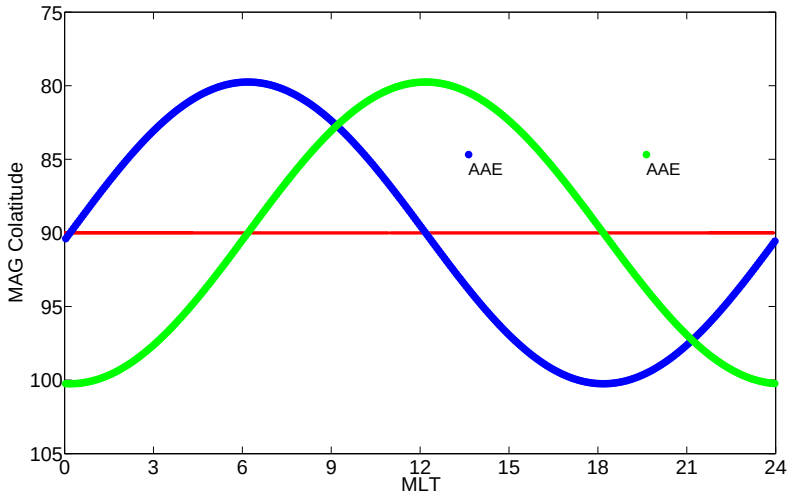

Figure 4.4: MAG-GEO equator phase in MLT for different start-UTs. The blue and green lines are series of coordinates along the GEO equator which have been converted to MAG coordinates, assuming they occurred at different UTs. The blue and green lines are then the position of the GEO equator in MAG coordinates for the start UTs of 11:00 (blue) and 17:00 (green). The red line is the MAG equator. The two isolated points are the location of the Addis Ababa (AAE) geomagnetic observatory in MLT coordinates (for reference only), coloured blue or green according to the UT used in the transformation from the original GEO coordinates.

given a 'viewpoint' at a certain LT.

Since the ecliptic oscillates about the GEO equator (cf. the Geocentric Solar Magnetospheric (GSM) coordinate system in Hapgood, 1992, where the x-axis is the Earth-Sun line) rather than the MAG equator, any seasonal signal obtained via an EOF analysis will then have a hemispheric bias dependent on the choice of start-UT. Selecting a single start-UT as the basis of my analysis is likely to bias the results. Running the EOF analyses separately on all 24 possible start-UTs and combining the 24 patterns for each EOF mode will ameliorate both the issues of station distribution and the GEO equator angle in MAG (which averages to zero if all start-UTs are combined). I will later demonstrate that the assumption that each EOF analysis is indeed representing the same signal in each start-UT (for at least mode 1) is valid.

Approaches to mitigating the effects of missing data in EOF analyses are discussed by Jolliffe (2002, section 13.6) and Beckers & Rixen (2003). The simplest approach is to replace the missing values with zeros – since I have centred the data matrix, zero value infilling is equiv-

alent to replacing the missing values with the mean of the data set (assuming the mean value I initially removed was an unbiased estimate). A more complex approach is to use the zero-infilling method to compute a set of EOFs, then use the first mode to reconstruct the missing values of the data matrix (using equation (4.7)). The new data matrix is then EOF-analysed a second time, producing a more accurate EOF solution.

In this study, I seek to retain the spatial signal of the daily variation, and use a degree 2, order 1 hour-by-hour SHA of the existing data to infill missing values. This approach is not perfect, but is considered less unstable than applying an EOF analysis to partially-complete data. Figure 4.5 show the percentage of missing (or infilled) observatory data in each year of the data set. Note the larger quantities of missing data in the year 2010.0-2011.0 (labelled 2010) – I avoid this year to limit error in my analyses. Note from Figure 4.2 that this year does not add significantly to the full solar cycle I am studying. EOFs cannot accept a varying station distribution, so after the missing-value interpolation, I treat the observatory record as a continuous time series.

My focus on the temporally-averaged character of the LT-fixed fields necessitates the use of additional data selection for magnetically quiet times only. These are selected from the internationally-defined five quietest days per month. If there were not data in each month, then this would disrupt an even sampling of the cyclostationarity of the field (Jolliffe, 2002). Cyclostationarity refers to the tendency of the field to repeat coherent spatial structure on a regular temporal basis (for instance, daily or annually). It is this property I take advantage of when selecting only one UT per day. The error introduced by an inconsistent selection of days within each month is considered offset by the guarantee that there is a uniform number of days' coverage in each month analysed. The use of the five quietest days per month is a measure of quietness which is relative to the mean activity at the time of the month in question, whereas the quiet-time data selection used in determining the lithospheric field observatory bias (section 4.6.1) was an absolute measure of quietness. The use of a relative field-activity selection method is required since an absolute measure of field activity would not be able to ensure a

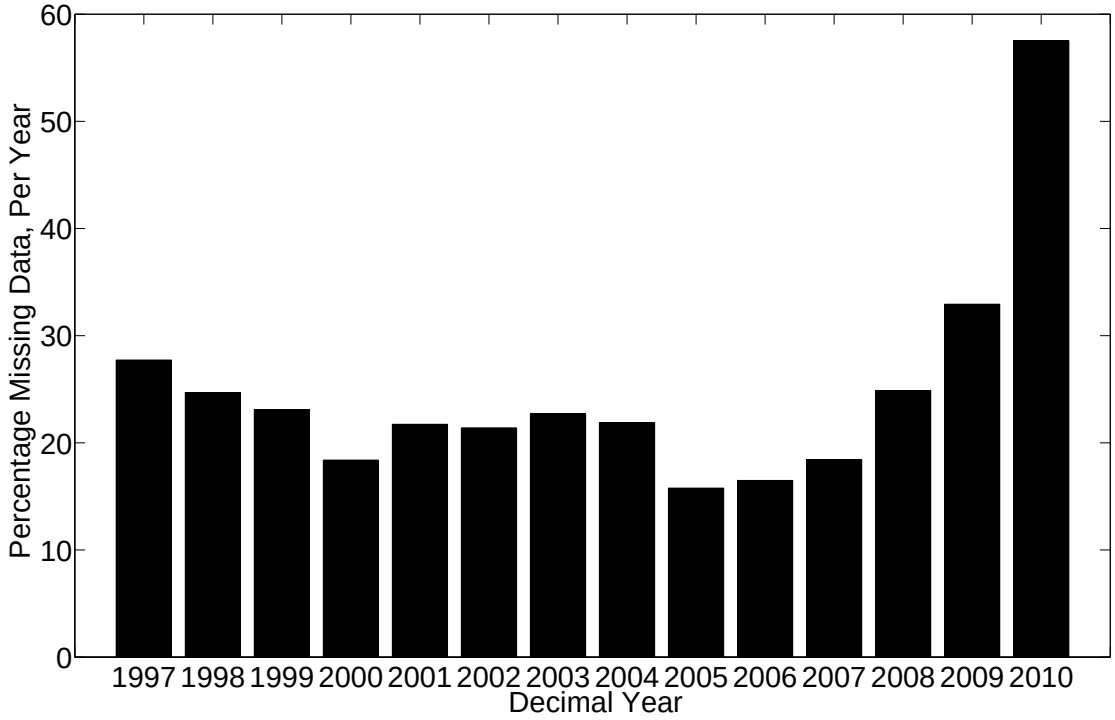

Figure 4.5: Percentage of missing data for all observatories in each year from start-1997 to end-2010.

consistent data coverage in time (I would not be guaranteed data coverage in each month). I therefore introduce more field activity at solar maximum, where it is more common. An investigation (the results of which are not shown) was carried out into the effect of field activity on the EOF analysis – the error introduced by the levels of field activity I retain is minimal. Note that the centring of the data matrix (equation (4.2)) is performed after the selection of only quiet days using the relative-activity method described here, which itself comes after the removal of the lithospheric field observatory bias from the entire observatory record. I therefore remove the quiet-day means twice, using different methods. This is overkill, but ensures that the centring of the data matrix is ultimately pertinent to the timespan of data being analysed.

4.6.2 Distribution weighting

The final modification I apply to my data before solving for the EOFs is a station-distribution weighting metric. EOFs do not take station distribution into account – since the functions

respond to shared variance, any spatial clustering of stations will act as a disproportionate attractor for the EOF projections. EOFs are defined in space and time, and it is possible to apply both spatial and temporal weighting metrics – this approach is discussed in detail by Baldwin et al. (2009), and Jolliffe (2002, page 382). I avoid temporal weighting as a relatively even temporal coverage has been ensured already, and apply only a set of spatial weights to mitigate the effects of the irregular station distribution. An example of spatial weighting given by Wilks (2005, page 479) describes the eigenanalysis of data gridded on an equal latitude-longitude mesh. In Wilks (2005)'s example, the variance projections are prevented from clustering at the geographic poles via a multiplication by the square root of the cosine of the latitude of each grid cell, to offset the effect of information duplicated between nearby cells. The square root is used because the weighting metric is squared when calculating the covariance matrix (the precise equation follows later in this section), and the cosine is applied to reflect the fact that the amount of independent information contained in each grid cell is not necessarily directly proportional to its area. Since the application of EOF analyses in climatology is typically preceded by a gridding routine, there are few predefined approaches for the more complex weighting distributions that I require. I have developed a method of quantifying the distance between a station and its nearest neighbours via a triangulation of the observatory distribution, achieved by calculating the convex hull (Weisstein, 2003) of the distribution of station locations in MAG coordinates.

To avoid edge effects, the first step in the triangulation is to replicate the global distribution of observatories, providing a margin of station locations which are consistent in an equal latitude-longitude grid (but which have no physical basis). This prevents the near-polar stations, and those at longitudes near 0 or 360°, from having larger/smaller triangulation areas than they should. Figure 4.6 shows a schematic of the global distribution replication. The numbers inside each rectangle are the extreme colatitudes and longitudes of the distribution. Note that the top and bottom layers of replicated distributions have been reversed in colatitude and shifted 180° in longitude, in order to make the observatory locations outside of the main coordinate system span commensurate with those inside it. Figure 4.7 shows the resulting tri-

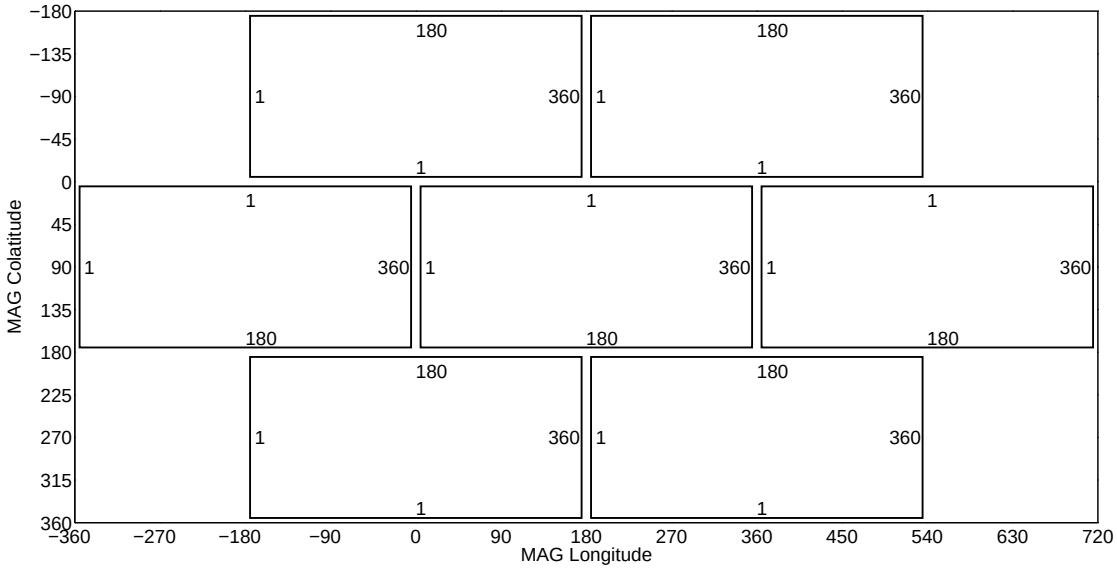

Figure 4.6: Schematic of global points distribution replication process, intended to avoid edge effects in the central rectangle's distribution triangulation (shown in Figure 4.7). The numbers in each rectangle are the extreme colatitudes and longitudes of the distribution.

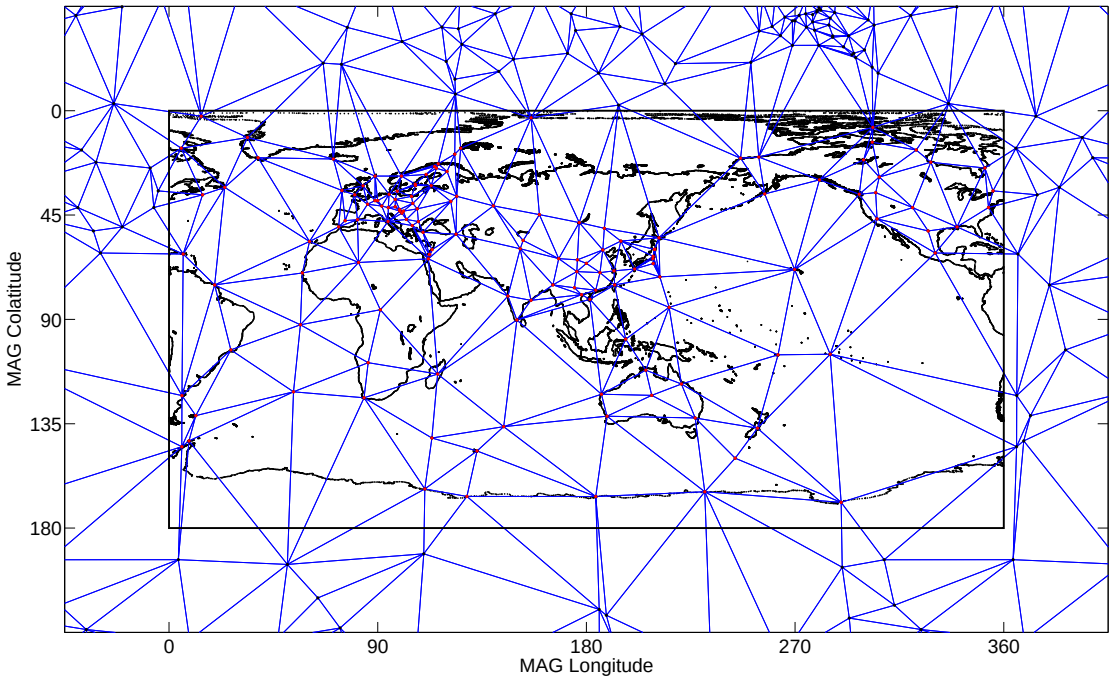

Figure 4.7: Triangulation of the replicated observatory distribution. Black vertices are non-physical, used to control edge-effects. Red vertices are the station locations I use. Blue lines indicate the spherical triangles forming the convex hull. MAG coordinates are used here.

angulation, performed in the spherical polar coordinate frame (GEO). The weighting metric is an estimation of the amount of 'free space' occupied by each observatory, and is given by the summation of the areas of each spherical triangle which has that observatory as a vertex. Calculation of the spherical triangle area is performed in the Cartesian frame, as follows.

In Figure 4.8, a schematic of a single spherical triangle between three observatories (red dots) is shown. The vectors **a**, **b** and **c** are each of length R_E (Earth's radius) and connect the centre of the Earth to each of the three observatory locations. Connecting the three observatories to each other are the vectors **d**, **e** and **f** – these lie along the surface of the sphere, and are shown in Figure 4.8 via their unit vectors $\hat{\mathbf{d}}$, $\hat{\mathbf{e}}$ and $\hat{\mathbf{f}}$. The unit vectors are calculated in Cartesian coordinates using

$$\begin{aligned}
\mathbf{d} &= \mathbf{a} \times (\mathbf{a} \times \mathbf{b}) \\
\mathbf{e} &= \mathbf{b} \times (\mathbf{b} \times \mathbf{c}) \\
\mathbf{f} &= \mathbf{c} \times (\mathbf{c} \times \mathbf{a}) \\
\hat{\mathbf{d}} &= \mathbf{d} / ||\mathbf{d}|| \\
\hat{\mathbf{e}} &= \mathbf{e} / ||\mathbf{e}|| \\
\hat{\mathbf{f}} &= \mathbf{f} / ||\mathbf{f}|| \,.
\end{aligned} \tag{4.8}$$

I seek the angles (on the surface of the sphere) g, h and i between the unit vectors $\hat{\mathbf{d}}$, $\hat{\mathbf{e}}$ and $\hat{\mathbf{f}}$. These angles are calculated (in Cartesian coordinates) via a simplification of the dot product given by Weisstein (2003)

$$\begin{aligned}
g &= \arccos\left(\hat{\mathbf{d}} \cdot -\hat{\mathbf{f}}\right) \\
h &= \arccos\left(-\hat{\mathbf{d}} \cdot \hat{\mathbf{e}}\right) \\
i &= \arccos\left(-\hat{\mathbf{e}} \cdot \hat{\mathbf{f}}\right) .
\end{aligned} \tag{4.9}$$

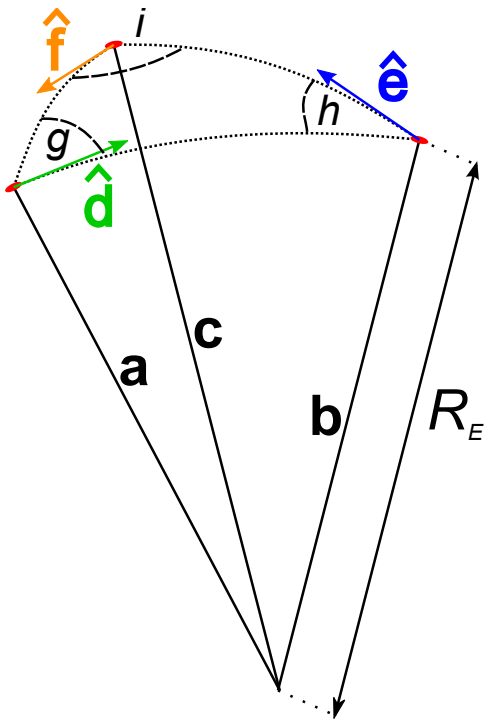

Figure 4.8: Schematic of spherical triangle area calculation. Vector labels are explained in the main text.

Girard's Spherical Excess formula (Weisstein, 2003) relates the angles of a spherical triangle to the 'spherical excess' E via

$$E = g + h + i + 180^\circ. \tag{4.10}$$

And the surface area (A) of the spherical triangle is then given by (Weisstein, 2003)

$$A = R_E^2 E, \tag{4.11}$$

where R_E is the radius of the sphere, in this case the Earth's radius approximated to 6371.2 km.

The vector of weighting metric values $\mathbf{w}$ is of length p, having one weighting value for each observatory. Each observatory's weighting value is formed from the summation of the surface areas A of each spherical triangle (in Figure 4.7) which surrounds that observatory, as

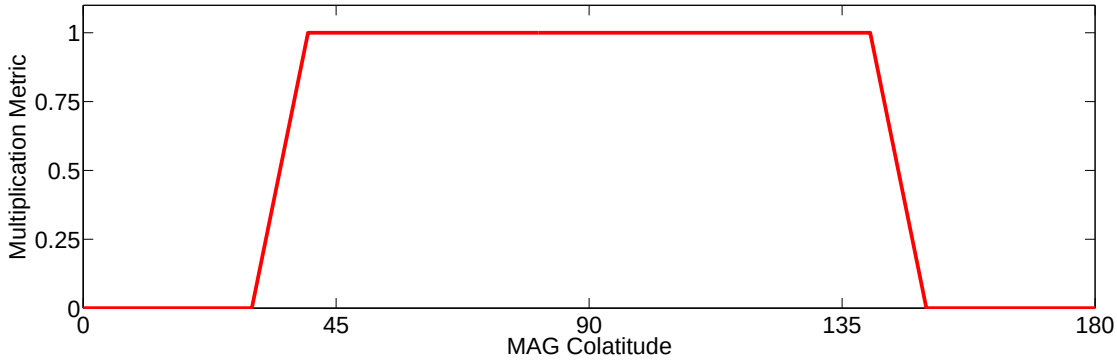

Figure 4.9: Polar-region mask based on MAG colatitude, designed to screen out magnetic field signal from polar and auroral regions. It is multiplied with the summed spherical triangle areas **w** (from Equation (4.12)) to calculate the final weighting metric.

follows

$$w_j = \sum_{k=1}^{n} A_{jk} \tag{4.12}$$

where n is the number of spherical triangles surrounding the j^{th} observatory. The metric values are retained only for the 142 (p) red points in Figure 4.7. I seek to remove the effects of the auroral regions from my analyses, and so multiply **w** (according to its associated observatory MAG colatitude coordinates) by a polar-region mask (with 10°-wide linear fade-out zone, calculated in MAG colatitude). The mask is shown as a function of MAG colatitude in Figure 4.9.

The weighting matrix **W** is a square matrix of dimension p, calculated via

$$\mathbf{W} = \mathbf{wI} \tag{4.13}$$

where **I** is the p x p identity matrix. The weighting matrix is applied to the data matrix prior to the EOF analysis, either by

$$\widetilde{\mathbf{X}}' = \widetilde{\mathbf{X}}\mathbf{W}^{1/2}, \tag{4.14}$$

or by

$$\mathbf{R}' = \left(\mathbf{W}^{1/2}\right)^{\mathsf{T}} \mathbf{R}\mathbf{W}^{1/2}. \tag{4.15}$$

Equations (4.14) and (4.15) will produce identical EOF expansions of the data. Note that the use of the square root of $\mathbf{W}$ is because the weighting matrix is squared in the process of calculating $\mathbf{R}$. The weighting matrix is intended to prevent clusters of stations (*e.g.* Europe) from dominating the partitioning of variance. However, in deriving the metric I have assumed that the amount of independent information contained in each observatory record is proportional to the area between it and its neighbours. This is likely not the case, as discussed by Jolliffe (2002, page 385), and in the grid-cell weighting example from Wilks (2005), discussed above. The observatories will record contributions from sources of widely ranging spatial scales. Jolliffe (2002) discusses the benefits of applying a further scaling to the derived weighting metrics – an exponent between 0.5 and 1, reflecting the duplicated information in even distantly-spaced observatory records. The application of a range of such scaling factors were tested (results not shown), but the effects were minimal on the leading modes, so no further scaling was applied to $\mathbf{W}$ (beyond the already-present exponent of 1/2).

Following Equation (4.15), the EOF expansion from Equation (4.4) is now (Baldwin et al., 2009)

$$\mathbf{R}'\mathbf{B}' = \mathbf{B}'\mathbf{L} \tag{4.16}$$

where $\mathbf{B}'$ are the weighted EOFs, which diagonalise $\mathbf{R}'$. The temporal coefficients ($\mathbf{Y}'$) of the weighted EOFs are given by

$$\mathbf{Y}' = \widetilde{\mathbf{X}}'\mathbf{B}' \tag{4.17}$$

Whilst the spatial weighting only affects the spatial layout of the EOFs (*i.e.* not their temporal progression), the orthogonality of the expansion applies only to the diagonalisation of the weighted data. Hence I have termed the PCs from the expansion of the weighted data $\mathbf{Y}'$,

to avoid confusion with the variable **T**. This change in terminology will also aid in later discussion, as the de-weighted **Y** (defined later in this section) is not equal to the un-weighted **T** because of the information lost in the application of the polar mask.

The application of the weighting metric has allowed me to compute an orthonormal basis expansion (the weighted EOFs) for the weighted data. For display purposes, and to interpret the expansions in the units of the pre-weighted data, I must extract a 'de-weighted' version of the data matrix, as follows (Jolliffe, 2002)

$$\widetilde{\mathbf{X}} = \widetilde{\mathbf{X}}'\mathbf{W}^{-1/2} \tag{4.18}$$

Equation (4.18) can also be used following any partial summation of $\widetilde{\mathbf{X}}'$ using the reconstruction equation (4.7), substituting the weighted-version PCs and EOFs as follows

$$\widetilde{\mathbf{X}}' = \sum_{i=1}^{p} \mathbf{y}'_i \mathbf{b}'^{\mathsf{T}}_i. \tag{4.19}$$

It is possible to compute the de-weighted EOFs via $\mathbf{B} = \mathbf{W}^{-1/2}\mathbf{B}'$ (Jolliffe, 2002), though I do not display these. Computing a de-weighted version of $\mathbf{Y}'$ for display purposes is more complex, since I cannot validly remove the spatial weighting matrix. The key lies in recognising that the units of the PCs are arbitrary, but consistent (hence multiplication of **T** with **V** produces **X**, as in Equation (4.6)). I can obtain **Y**, which is a version of $\mathbf{Y}'$ corresponding to the range and units of $\widetilde{\mathbf{X}}$, via the scaling operation

$$\mathbf{y}_j = \mathbf{y}'_j \left| B_{ji} \right|, \tag{4.20}$$

where $\mathbf{y}_j$ is the j^{th} columnar vector of **Y**, j is from 1 to p (recall that there are p EOFs and p PCs), and $\left|B_{ji}\right| = \left|\mathbf{b}_j\right|_{max}$ (with $\mathbf{b}_j$ the j^{th} columnar vector of **B**). Note that von Storch &

Zwiers (2002, page 298) described a 'neater' definition of the attribution of data units to the EOF expansion, but one which does not account for the application of weighting metrics to the input data. I plot the de-weighted versions $\mathbf{Y}$ and $\widetilde{\mathbf{X}}$.

4.7 Synthetic data tests, and how to display the EOF output

Later discussion and analysis will benefit from a demonstration of the methods used to represent the output of an EOF analysis. In this section I use a synthetic-data EOF example to give physical meaning to the expected EOF projection-attractors discussed in section 4.4. Figure 4.10 shows the input data for this example, which is the ionospheric and ionospheric-induced signal data prediction of the CM4 model (Sabaka et al., 2004). I use the ionospheric data prediction as a test-case because it has a seasonal oscillation as well as asymmetry in MLT. I include the induced signal to add spatial complexity. The prediction is made at 24/hourly intervals (with a UT of 18) from 1997.0 to 1998.0, onto a 648-node equal-area[3] tesseral partitioning of the sphere using algorithms developed by Leopardi (2007)[4]. I hereupon refer to these data as the synthetic data set. Figure 4.10(a) and (c) show the data from the θ- and ϕ-components (respectively) in northern hemisphere winter, whilst panels (b) and (d) show the same two components, but in northern hemisphere summer.

The EOF analysis of the synthetic data is performed without spatial weighting (the equal-area grid is optimal with regards to spacing), though a 20°-wide MAG colatitude polar-cap mask is applied to the input data to avoid high-latitude effects (should any arise). I introduce the following series of figures to establish familiarity with the method of displaying the EOF analysis output, in a controlled setting. The percentage of the variance of the input data explained by each mode is calculated with $100(L_0/\sum_{i=1}^{n} L_i)$, where n is the total number of modes, and L_0 is a certain eigenvalue (Bjornsson & Venegas, 1997; Jolliffe, 2002).

[3]Note that whilst there are no partitions shown in Figure 4.10, and thus, the term 'equal-area' would seem to lack meaning, the points shown are the centroids of an equal-area partitioning of the sphere, and are considered an optimal data layout in this synthetic example.

[4]Matlab code to compute the partitioning of the sphere was supplied by Ciaran Beggan, 2009, originally obtained from Paul Leopardi.

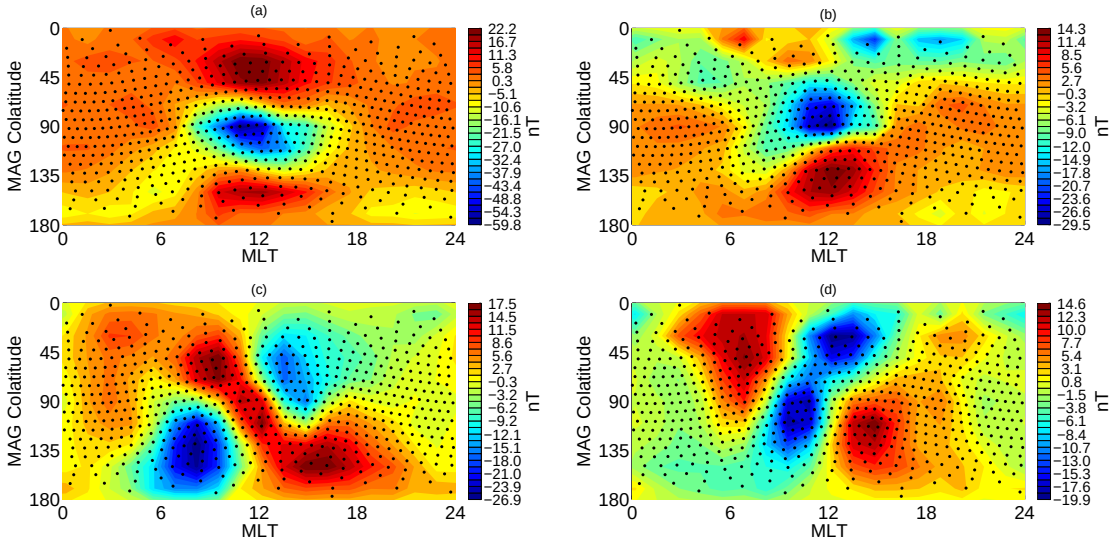

Figure 4.10: CM4 model ionospheric and ionospheric-induced signal data prediction on a 648-node equal-area tesseral grid. Panels (a) and (b) are for the θ-component, whilst panels (c) and (d) are for the ϕ-component. Panels (a) and (c) show northern hemisphere winter, whilst panels (b) and (d) show northern hemisphere summer. Black dots are the data prediction locations, on a sphere of radius analogous to the Earth's surface (6371.2 km). The longitudinal curvature of these points is due to their calculation in GEO before transformation to MAG.

Figures 4.11 and 4.12 show the θ- and ϕ-component eigenspectra, respectively, as a function of the percentage variance explained by each mode. Both show good separation of variance, with the first mode accounting for over 75% of the variance in each component. The eigenspectrum is analogous to the power spectrum (Lowes, 1974) since both describe the distribution of variance across the basis coefficients (von Storch & Zwiers, 2002, page 303), though that is where the similarities end. Were any two EOFs to have identical eigenvalues, then any linear combination of their patterns would be equally relevant (given the constraints of the expansion), and that pair of EOFs would be termed 'mathematically degenerate' (von Storch & Zwiers, 2002; Hannachi et al., 2007, pages 295, 1123 respectively). As discussed earlier, this is why I avoid travelling waves in the EOF input. However, the real case is more complex than looking for identical eigenvalues. Since the data matrix is a sample population of the underlying functions I am measuring, then the EOFs I compute are in fact sample estimates of the 'true' EOFs (von Storch & Zwiers, 2002, page 299). The eigenvalues therefore also have sampling errors. As stated by North et al. (1982, page 704) "if the sampling error of a particular eigenvalue is comparable to or larger than the spacing between it and a neighboring

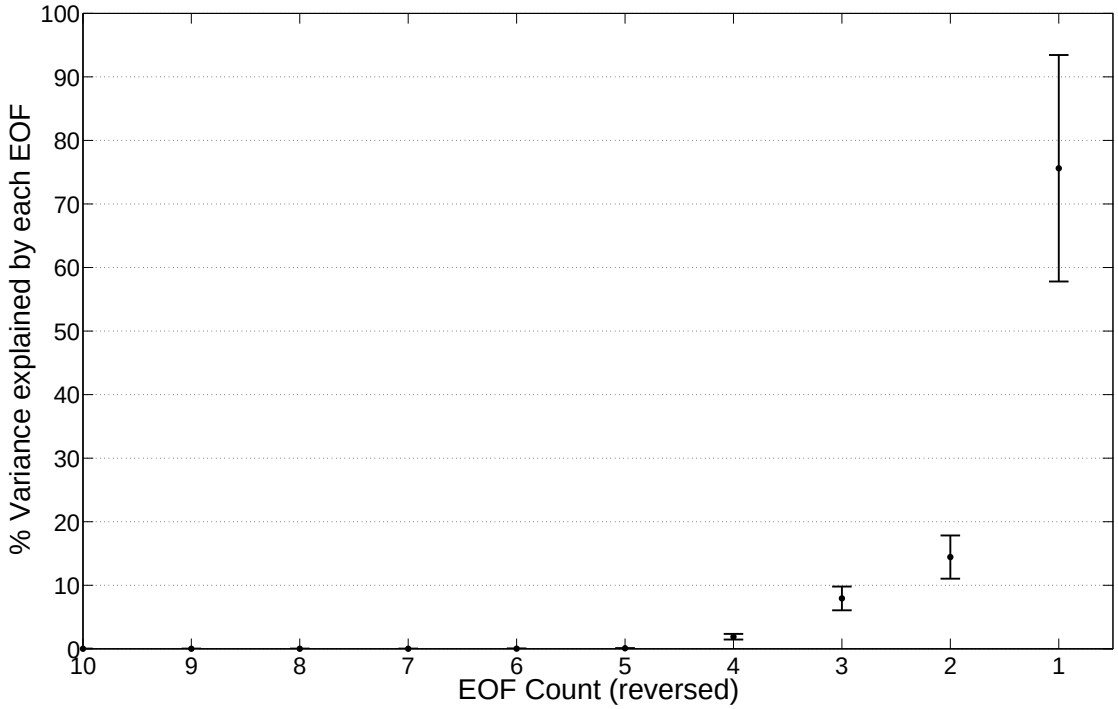

Figure 4.11: Eigenspectrum of θ-component EOF analysis of synthetic ionospheric data. Error bars are given by North's rule of thumb (explained in main text). Only 10 of the full 648 EOF modes are shown, to better delineate the order of the leading modes.

eigenvalue, then the sampling errors for the EOF associated with that eigenvalue will be comparable to the size of the neighboring EOF". Any EOFs within this range can be considered mathematically degenerate, and not representative of any physically meaningful pattern. The error bars on Figures 4.11 and 4.12 describe a measure of the mathematical degeneracy of the EOF patterns, called 'North et al.'s rule of thumb' (North et al., 1982; Wilks, 2005, pages 704, 489 respectively) given by;

$$e_L = \mathbf{L}\left(\sqrt{\frac{2}{D}}\right), \tag{4.21}$$

where e_L is the eigenvalue sampling error (half the length of the errorbars in Figures 4.11 and 4.12), and D is the number of independent data in the sample (which I approximate with an estimate of the degrees of freedom of the data matrix).

The estimation of mathematical degeneracy is a useful parameter, important as the first step

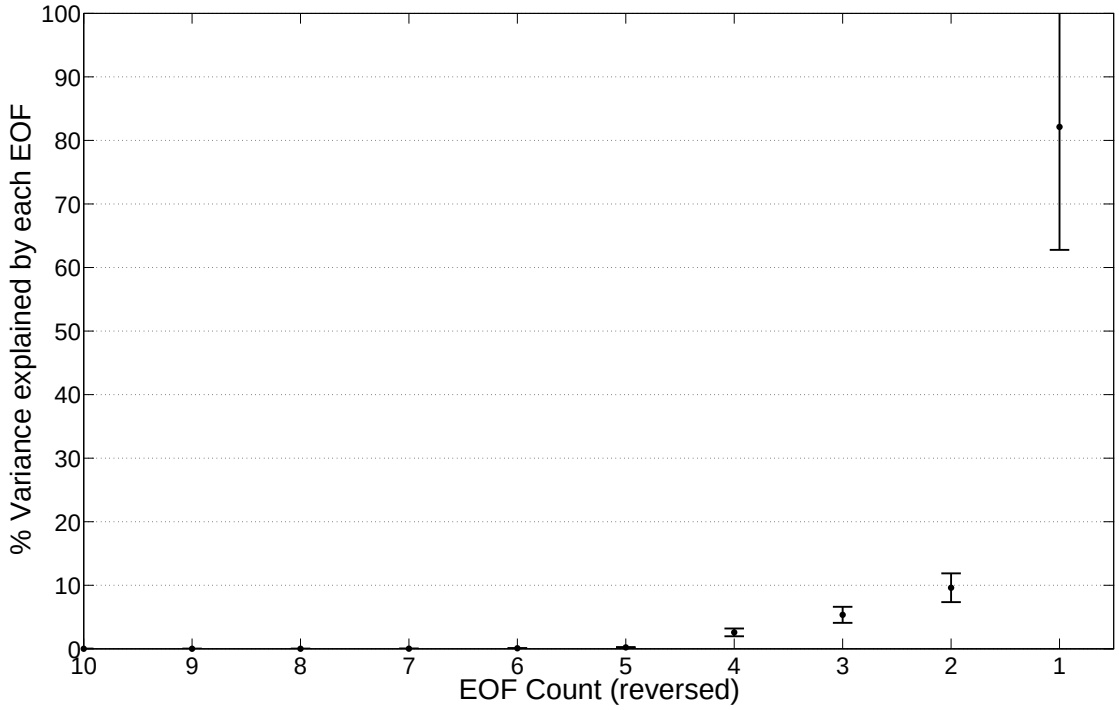

Figure 4.12: Eigenspectrum of ϕ-component EOF analysis of synthetic ionospheric data. Error bars are given by North's rule of thumb (explained in main text). Only 10 of the full 648 EOF modes are shown, to better delineate the order of the leading modes.

in assessing the possible physical significance of EOFs – I will not assign physical significance to any EOF which is considered degenerate – but the use of this metric has severe limitations, as follows. In real data it is common for the non-leading modes to exhibit significant degeneracy before the eigenspectrum fully flattens. It would be tempting to assume that the mode at which mathematical degeneracy begins to dominate the eigenspectrum also denotes the cut-off of the signal-subspace. However, as stated by von Storch & Zwiers (2002, page 303), the shape of the eigenspectrum is not directly linked to the dynamical content of the underlying function. This implies that the degenerate parts of the eigenspectrum may contain otherwise-useful signal in the form of 'entangled' EOFs which cannot be separated. Put another way, although I cannot define the signal-subspace from the eigenspectrum alone, the sampling errors may force me to place the signal-subspace cut-off closer to the leading mode than I would have otherwise. Furthermore, the shape of the eigenspectrum does not help to distinguish between physical sources (Monahan et al., 2009).

To determine the physical significance of the modes, I rely upon the EOF data reconstructions and the PCs. Figures 4.13 and 4.14 show the θ- and ϕ-component synthetic data matrix reconstructions for the four leading EOFs (separately), each using Equation (4.7). The reconstructions were performed at the timestamps of maximum amplitude of each EOF's PC, which makes the range of the gridded reconstruction equal to the amplitude range of its PC. The θ- and ϕ-component PCs are shown in Figures 4.15 and 4.16 respectively. As expected from discussion in section 4.4, the first PC of both the θ- and ϕ-components is the long-period annual (seasonal) oscillation.

Note that the θ-component synthetic data in Figure 4.10(a) are maximal at 12 MLT and 90° MAG colatitude, with two lower-magnitude peaks to the north and south, at the same LT. A similar arrangement is seen in panel (b), though the central peak now spreads more northwards than southwards, due to the change in season. The pattern of θ's EOF 1 in Figure 4.13 is commensurate with this seasonal oscillation – it consists of a low at the GEO equator (slightly north of the MAG equator at a UT of 18) and peaks to both the north and south, at 12 MLT, with some high-latitude effects in the northern hemisphere. The oscillation of PC 1 in Figure 4.15 will control the contribution of the θ-component's first EOF, so at two points in the year, there will be no input to the data matrix reconstruction from θ's EOF 1. Part of the annual signal must therefore be made up with the second θ-component EOF. The pattern of θ's EOF 2 has biannual PC oscillation, partly to replace the lost contribution from the zero-points of PC 1, and partly because of the increased particle influx into the ionosphere at equinox. EOF 2 has a tri-lobed structure with the same basic layout as the input data shown in Figure 4.10(a) and (b), though the northern hemisphere's lobe is muted, possibly because of the increased variance in the northern hemisphere in EOF 1. EOF 2 likely represents the day-to-day perturbations on the input data's spatial peaks.

Figure 4.10(c) and (d) shows the seasonal variation common to the ϕ-component, which is more spatially complex than that of the θ-component, exhibiting changes in both the sign and the structure of its patterns between summer and winter. The tendency of the ϕ-component

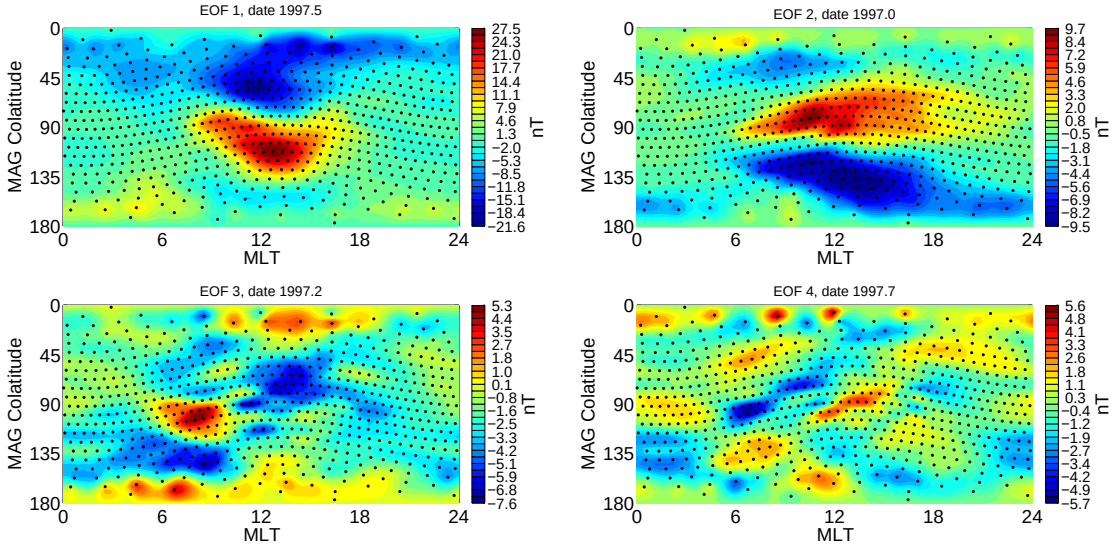

Figure 4.13: Reconstruction of synthetic θ-component data matrix for each of the first four EOFs, separately, using Equation (4.7). Black points indicate data locations. The reconstruction was performed at the timestamp of the strongest part of each mode's PC in Figure 4.15.

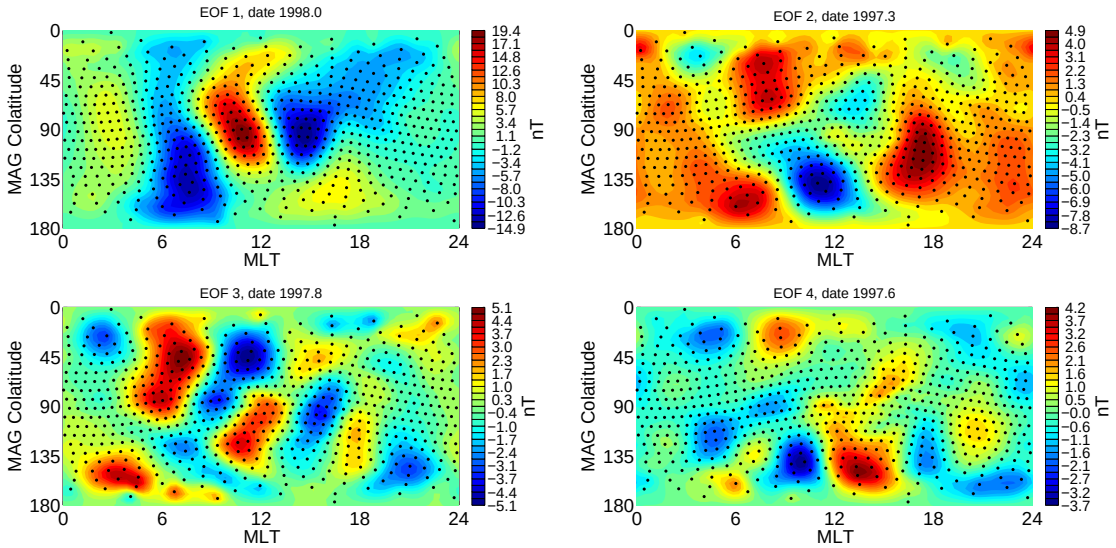

Figure 4.14: Reconstruction of synthetic ϕ-component data matrix for each of the first four EOFs, separately, using Equation (4.7). Black points indicate data locations. The reconstruction was performed at the timestamp of the strongest part of each mode's PC in Figure 4.16.

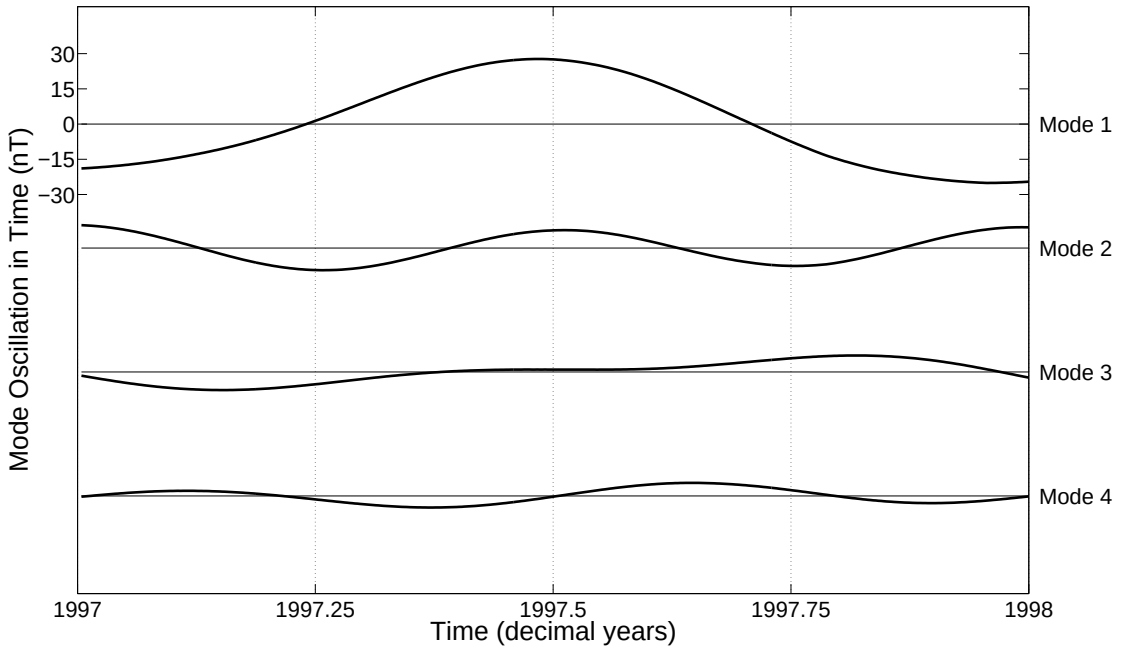

Figure 4.15: θ-component PCs from synthetic data EOF analysis. The first four modes are shown, illustrating the temporal progression of the spatially-static patterns shown in Figure 4.13.

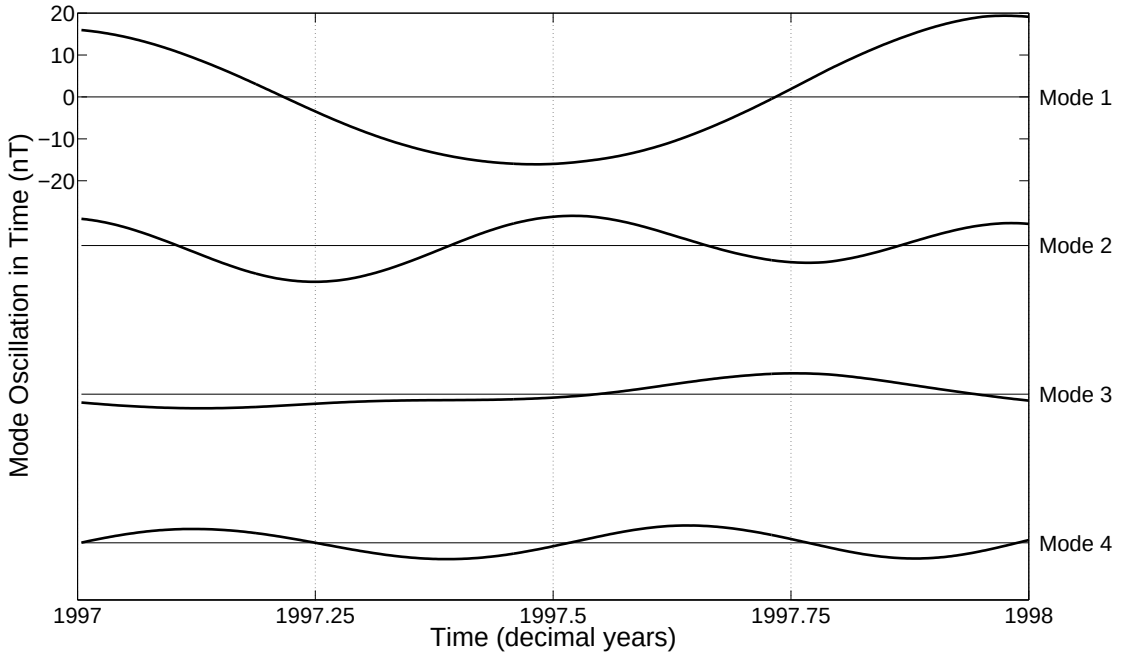

Figure 4.16: ϕ-component PCs from synthetic data EOF analysis. The first four modes are shown, illustrating the temporal progression of the spatially-static patterns shown in Figure 4.14.

signal to be more spatially variable than the θ-component is likely due to its lesser magnitude. The pattern of ϕ's EOF 1 in Figure 4.14 is commensurate with its seasonal oscillation, though this is more complex than for θ. EOF 1 has a longitudinal oscillation with season, though the sign changes of the input data make it difficult to specify if the peaks of EOF 1 are due to spatial changes or sign changes. The same uncertainty affects ϕ's EOF 2, though it is likely that (like for θ), this mode represents the day-to-day perturbations on the input data's spatial peaks. For both the θ- and ϕ-components, EOFs 3+ have low reconstruction magnitudes, spatial patterns which do not appear representative of any known physical current system of oscillation thereof, and similarly incoherent temporal oscillations. However, it is notable that mode 4 has a somewhat similar temporal progression to mode 2, indicating it to be a minor orthogonal part of the same dynamical aspect of the system, or high spatial frequency noise associated with mode 2's biannual temporal frequency. I therefore consider EOFs 3+ to represent the noise subspace. The real data will be more complex, and I will not be able to link the layout of the input data to the derived EOFs as easily as has been done here. The patterns of the θ-component EOFs are easier to link to known physical causes than those of ϕ, and I will focus on this component for much of the real-data analysis.

4.8 Results

4.8.1 Real data, single start-UT

Section 4.7 showed the results of an EOF analysis using an idealised data distribution. Here I show the results of an EOF analysis of real θ-component data, at a single start-UT of 18. From the eigenspectrum of this analysis, shown in Figure 4.17, it can be seen that the only non-degenerate mode is that of EOF 1, precluding a physical interpretation of the non-leading modes. The first four PCs of this analysis are shown in Figure 4.18. Note that the non-physical nature and mathemetical degeneracy of the non-leading modes (as well as their lesser contribution to the variance of the data matrix) does not mean that they are lower-magnitude than the leading mode. Mode 1 has an annual oscillation, and is the long-period signal I am attempting to isolate in the data. The non-leading modes do not appear to have a coherent temporal pattern.

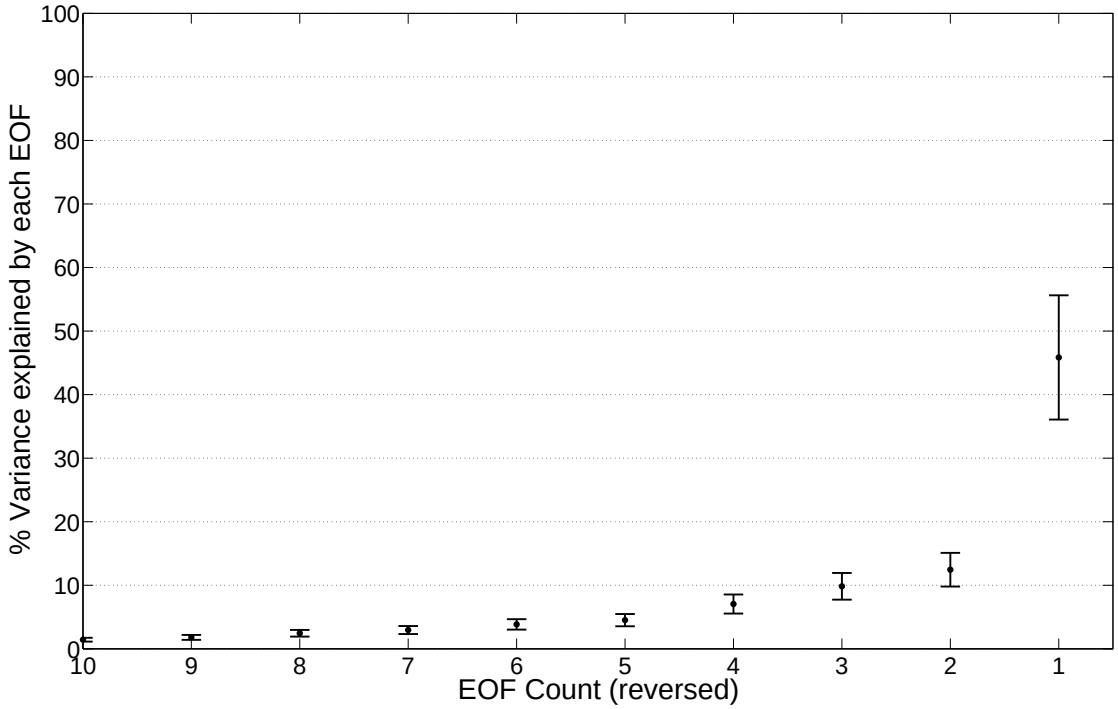

Figure 4.17: Eigenspectrum for an EOF analysis of θ-component data from 1997.0-1998.0, at a UT of 18. Only 10 of the full 142 EOF modes are shown, to better delineate the order of the leading modes.

Figure 4.19 shows the reconstructed data distributions for the first four modes. EOFs 2+ appear noisy, with high-magnitude values on the dayside but with no other coherent global pattern. EOF 1 is dominated by a north-south hemispherical-split – this is the seasonal oscillation apparent in the ionospheric synthetic data (in section 4.7). Laterally-speaking, the pattern of EOF 1 is strongest on the dayside, but is also present on the nightside. The continuation of the hemispherical-split signal across all MLTs may be due to ionospheric sources, or it may be due to the seasonal oscillation (in MAG) of the night-sector magnetospheric sources (such as the magnetotail) – we cannot tell from these data alone. The nightside section of the hemispherical-split signal is stronger on the dusk side (the region of interest here) of midnight than it is on the dawn side of midnight. If this dawn-dusk asymmetry signal is real, then we are likely seeing the effect of the partial ring current adding to the spatial pattern of the long-period oscillation, making it stronger over dusk. However, it is clear that the distribution of the stations is affecting the derived EOFs, despite the applied weighting. The start-UT of 18 is

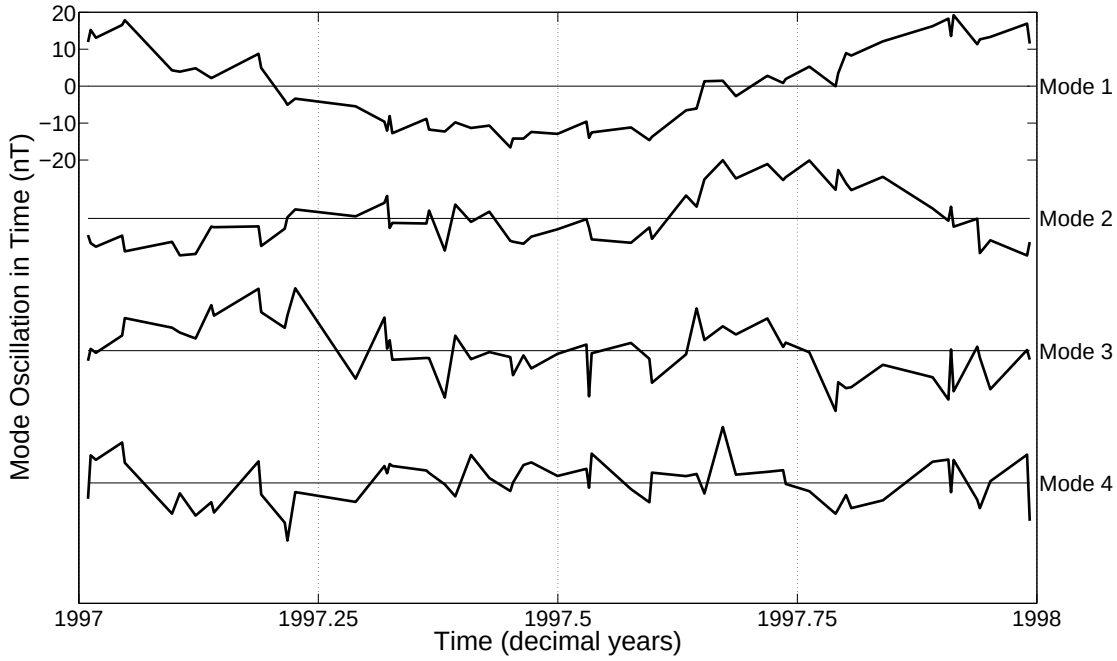

Figure 4.18: First four PCs for an EOF analysis of θ-component data from 1997.0-1998.0, at a UT of 18. The irregular spacing of the analysis points in time is due to the irregular separation of the quiet days used in the analysis.

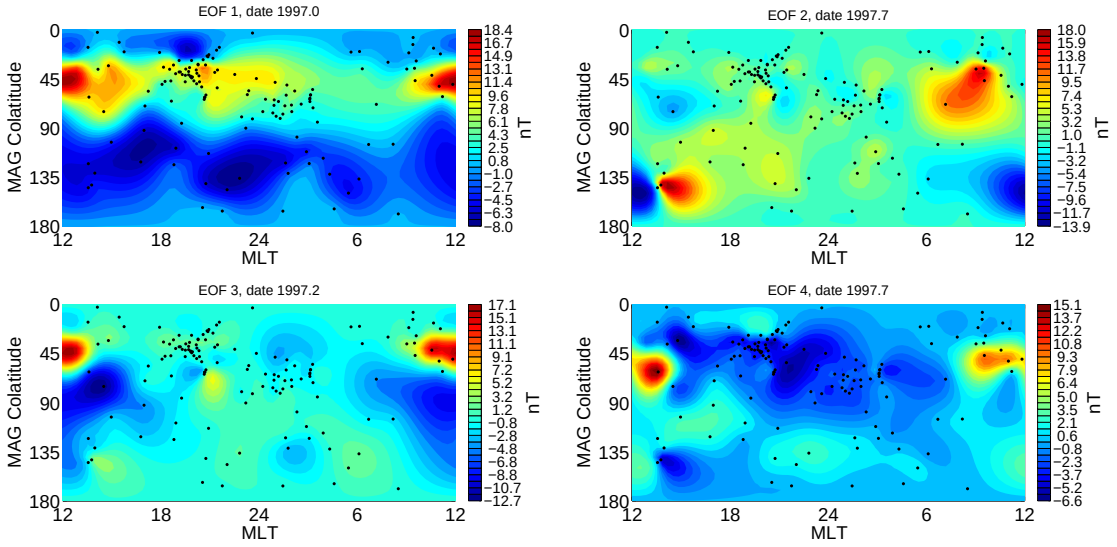

Figure 4.19: Data reconstructions for first four modes of an EOF analysis of θ-component data from 1997.0-1998.0, at a UT of 18. Black points are station locations. Reconstruction was made at strongest PC element for each mode. Note the shift in the centre MLT of the plot from 12 (as previously) to 24, which is used in all similar charts for the remainder of this chapter, to reflect the focus on the nightside.

used here because the MAG-GEO equator offset is nearly zero at dusk (refer to Figure 4.4, and because the densest clustering of stations does not occur on dayside (which always has high variance). The original intention for these analyses was to select an optimal station-distribution over the region of interest and use that as the analysis set-up. However, the station distribution effects are severe enough to warrant an approach which combines all start-UTs, as discussed in the following subsection.

4.8.2 Combining start-UTs

The distribution of the real data is not ideal, so I am required to apply certain post-EOF processing methods to make full use of the real-data EOF output. I discussed the basics of the approach of combining the 24 different start-UT EOF analyses in section 4.3. Combining the start-UTs of separate EOF analyses has the following benefits.

- Reduces bias from station distribution effects.
- The MAG-GEO equator effects should cancel out over the 24 different station-MLT location combinations.
- All the input data is used, rather than only $1/24^{th}$ of it, whilst avoiding travelling waves.

However, to combine the results of different EOF analyses, I must ensure that the following conditions are met.

- I must be certain that the reconstructions have the same sign (EOFs are sign-degenerate, as discussed in von Storch & Zwiers (2002, page 296)).
- The reconstructions must have comparable magnitudes if they describe the same point in the mode's temporal oscillation.
- The mode that I am combining the analyses for must represent the same field structure for each start-UT.

In certain year-long span EOF analyses throughout the solar cycle, the sampling error of mode 2 is quite distinct from mode 3 (not shown). However, modes 2+ are degenerate often enough

to warrant only focusing on combining start-UTs for mode 1 in the case of a year-long EOF analysis. As will be shown in a later subsection, an analysis of the length of the solar cycle has a much better signal-to-noise ratio, and hence, fewer degenerate modes.

For each year analysed, the leading mode of the (RC-corrected) θ-component is an annual (seasonal) oscillation, regardless of start-UT. With that established, a strategy for controlling the sign and magnitude of the reconstruction was developed, which starts with a data reconstruction for each start-UT's EOF analysis, at a single time. Figure 4.20 shows the first PC from Figure 4.18. The annual oscillation has been fitted with an order-6 polynomial (although lower-order polynomials would do an equivalently good job, an order-6 polynomial is required for the fitting shown in Figure 4.31, and I use the same order to fit the annual oscillation here to retain consistency). This allows selection of the maximum amplitude of the oscillation without the selection process 'snapping' to one of the high-frequency spikes. It is the amplitude of the polynomial expression, rather than the PC itself, which is used as $\mathbf{b}'_i$ in Equation (4.19) when making reconstructions for combining start-UTs. This enforces a more commensurate set of amplitudes for the combination of the data reconstructions. Note that the central section of the polynomial line in Figure 4.20 is coloured red – this indicates the segment of the oscillation which represents (northern hemisphere) summer. If I pick the maximum amplitude of each start-UT's mode 1 within this time span, the reconstructions will always have the same spatial distribution of sign. The data reconstruction occupies the length of the input time span even though only one mode is used for the reconstruction. I extract, from the full reconstruction, a set of values (at the observatory locations) corresponding to the time of the northern hemisphere summer-maximum in the oscillation (shown as a vertical green line in Figure 4.20). This was repeated for each start-UT, and the resulting distribution fed into a single-component SHA up to degree 5 and order 5, with reference radius 6371.2 km. In this process I treated the distribution of reconstructions as a scalar function, rather than as a component of a potential field.

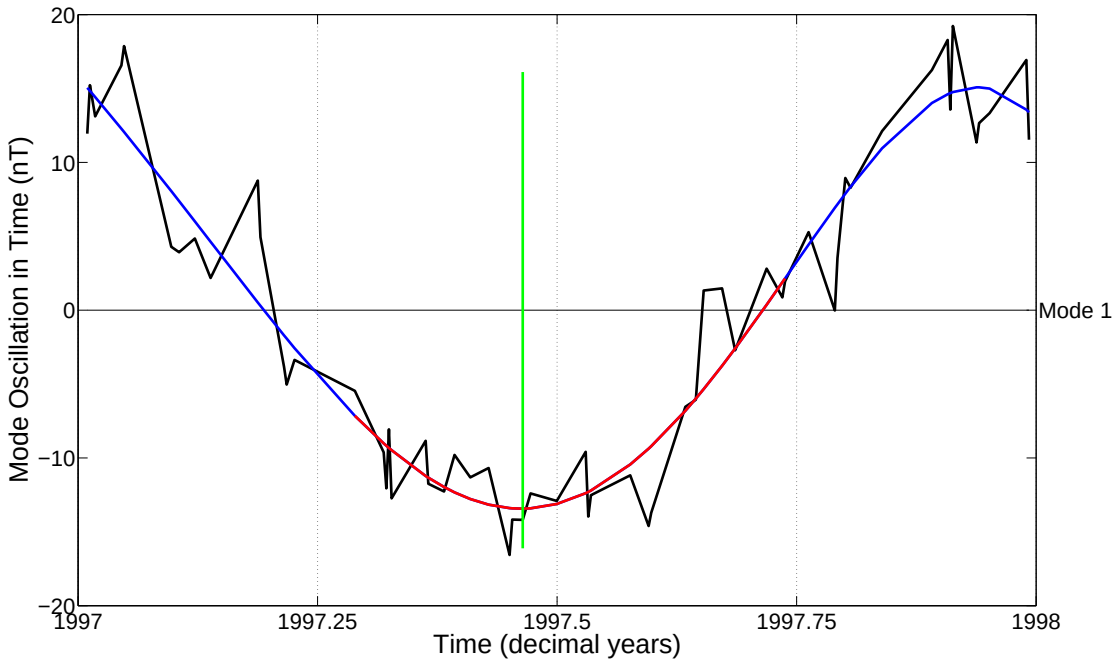

Figure 4.20: PC 1 (black line) for an EOF analysis of θ-component data from 1997.0-1998.0, at a UT of 18. The PC has been fitted with an order-6 polynomial (blue line) to reduce uncertainty in selecting the time and amplitude of the peak strength of the oscillation within northern hemisphere summer (shown in red). The time of the maximum amplitude is denoted by the green vertical line, and this controls the sign of the associated data reconstruction.

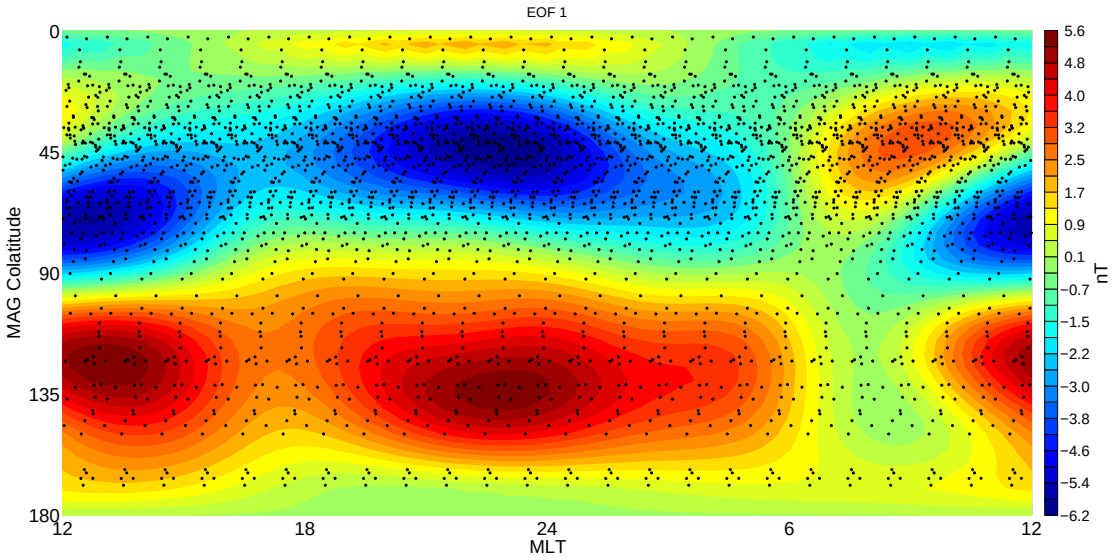

Figure 4.21: Data reconstructions of 24 θ-component EOF analyses for different start-UTs in the period 1997.0-1998.0, combined via SH analysis at degree and order 5. Black points are the distribution of input data for the SHA, derived from the 24-times replicated station distribution in LT (one distribution in LT for each of the 24 different start-UTs). Grid cell size is 5 degrees.

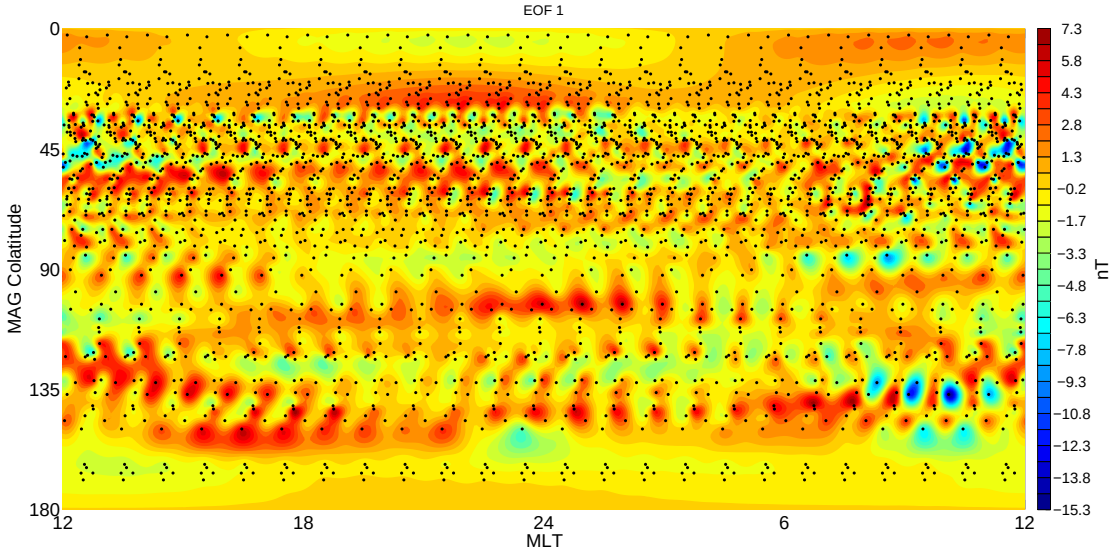

Figure 4.22: SHA data prediction residual from SHA of data reconstructions of 24 θ-component EOF analyses for different start-UTs in the period 1997.0-1998.0. Black points are the distribution of input data for the SHA, derived from the 24-times replicated station distribution in LT (one distribution in LT for each of the 24 different start-UTs). Grid cell size is 1 degree.

Figure 4.21 shows the results of combining 24 EOF analyses with different start-UTs from the time period 1997.0-1998.0, via the SHA data prediction. The SHA was performed on the replicated LT-distribution of all stations. The pattern shown in Figure 4.21 is the spatial signal of the long-period oscillation of the magnetic field's θ-component. It can be seen that the stronger signal at dusk with regards to dawn (identified in Figure 4.19) is indeed a real effect, and not due to station distribution. The magnitude range of the combined distribution is somewhat less than that of EOF 1 in Figure 4.19, since I have removed the higher-frequency noise – the SHA model will not fit this aspect of the data. Residuals to the SHA model are shown in Figure 4.22 – these lack globally coherent structure, but are typically stronger on the dayside.

4.8.3 Change in the annual oscillation spatial signal throughout the solar cycle

In this subsection I expand upon the single-year example distribution shown in section 4.8.2, and illustrate the change in the spatial distribution of the long-period signal at other years throughout the solar cycle. Figure 4.21 showed analysis from the start of the data coverage:

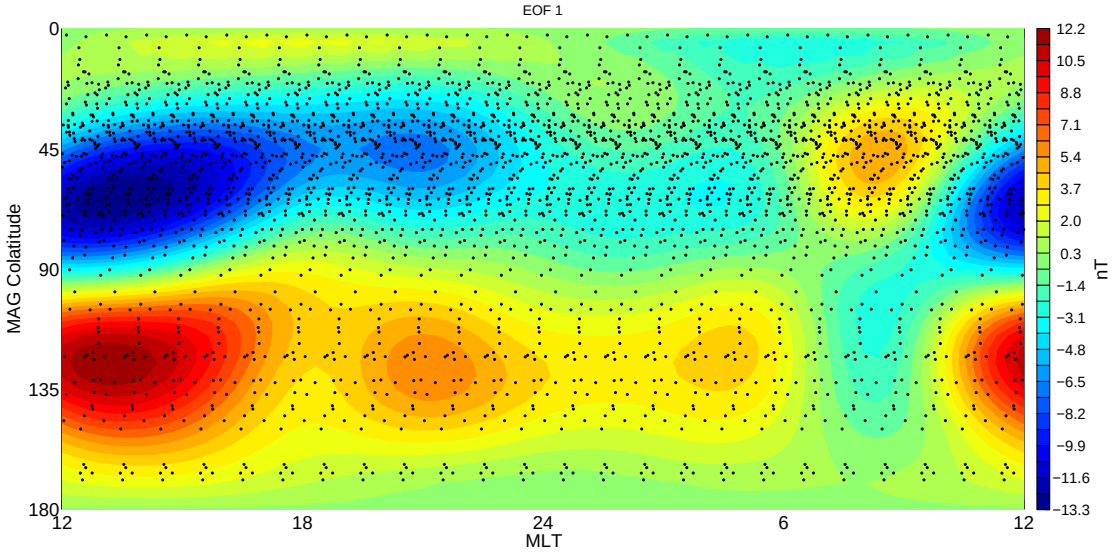

Figure 4.23: Data reconstructions of 24 θ-component EOF analyses for different start-UTs in the period 2000.0-2001.0, combined via SH analysis at degree and order 5. Black points are the distribution of input data for the SHA, derived from the 24-times replicated station distribution in LT (one distribution in LT for each of the 24 different start-UTs). Grid cell size is 5 degrees.

this is at a point of medium solar-cycle activity (refer to Figure 4.2). Figure 4.23 shows the annual signal from 2000.0-2001.0, produced in the same manner as described in section 4.8.2, representing a peak in the solar cycle. The $F_{10.7}$ values are actually higher (in terms of the peak amplitudes of transient events) in 2003-2004, but I have chosen 2000-2001 as the year of highest $F_{10.7}$ which is free of strong gradients in the solar flux density trend on long periods. In Figure 4.23, the strength of the pattern on dayside has increased due to higher solar activity affecting the ionosphere, making the nightside asymmetry appear smaller in magnitude (though it is approximately equal-magnitude to the nightside peaks in Figure 4.21), and the nightside asymmetry peak location has shifted slightly towards dusk. This duskwards-shift could indicate a higher contribution of the partial ring-current to nightside processes near the peak of the solar-cycle, or a LT-sector-shift in the maximum amplitude of the partial ring current. The latter possibility would be commensurate with the findings of Le et al. (2004) (that the peak effect of the partial ring current is between dusk and midnight), but a combination of both factors is likely.

To represent the trough of the solar cycle I use the span 2007.0-2008.0. The years after this

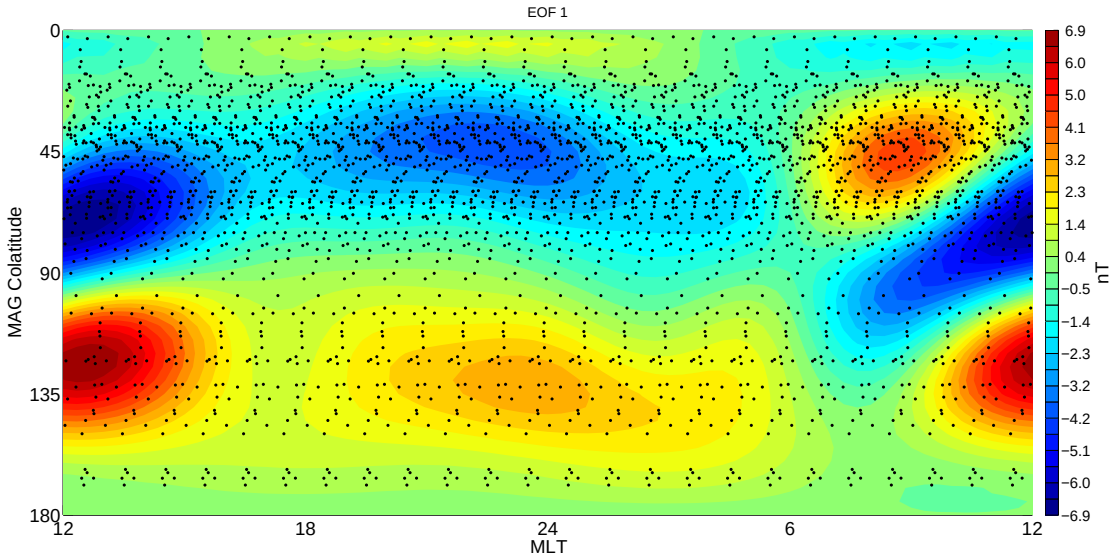

Figure 4.24: Data reconstructions of 24 θ-component EOF analyses for different start-UTs in the period 2007.0-2008.0, combined via SH analysis at degree and order 5. Black points are the distribution of input data for the SHA, derived from the 24-times replicated station distribution in LT (one distribution in LT for each of the 24 different start-UTs). Grid cell size is 5 degrees.

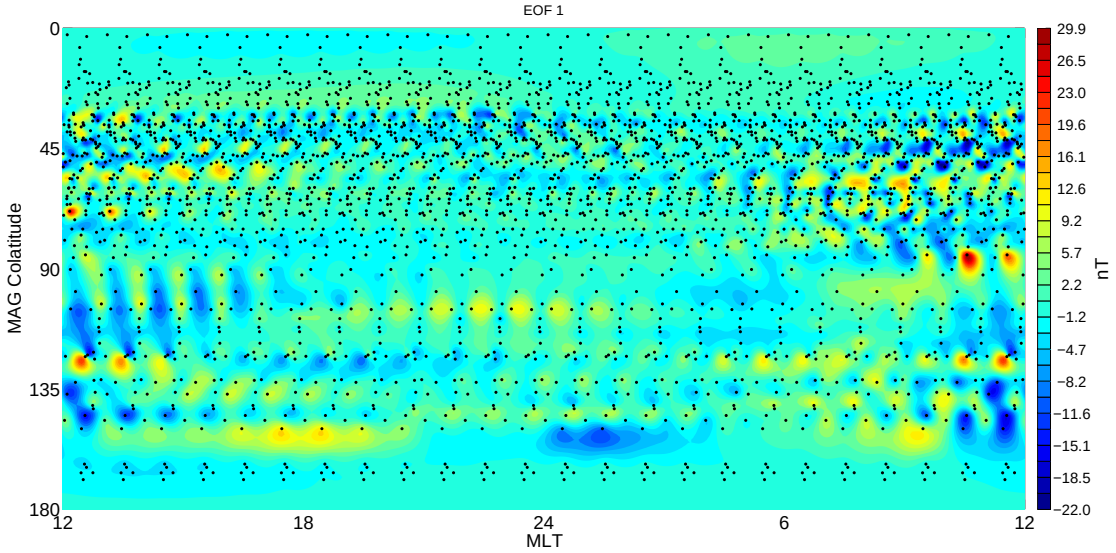

Figure 4.25: SHA data prediction residual from SHA of data reconstructions of 24 θ-component EOF analyses for different start-UTs in the period 2000.0-2001.0. Black points are the distribution of input data for the SHA, derived from the 24-times replicated station distribution in LT (one distribution in LT for each of the 24 different start-UTs). Grid cell size is 1 degree.

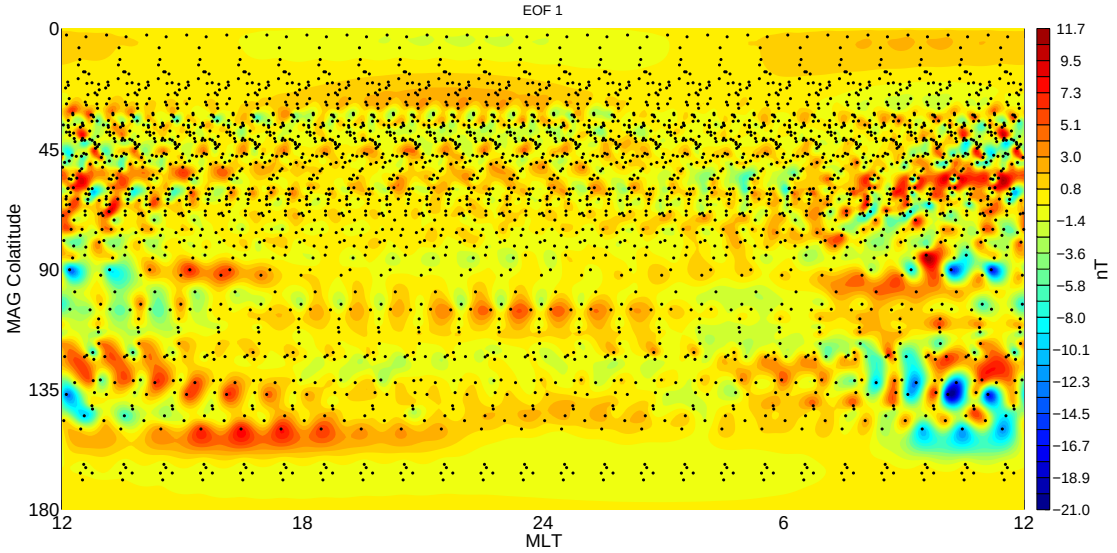

Figure 4.26: SHA data prediction residual from SHA of data reconstructions of 24 θ-component EOF analyses for different start-UTs in the period 2007.0-2008.0. Black points are the distribution of input data for the SHA, derived from the 24-times replicated station distribution in LT (one distribution in LT for each of the 24 different start-UTs). Grid cell size is 1 degree.

in the data coverage do have lower $F_{10.7}$ values, but (as shown in Figure 4.5), they have poorer data coverage. Figure 4.24 shows the annual signal from 2007.0-2008.0. In this year the nightside signal is lower-magnitude than either of 1997.0-1998.0 (Figure 4.21) and 2000.0-2001.0 (Figure 4.23), and the LT position of the nightside peak has shifted back towards midnight, occupying the same position as in 1997.0-1998.0. Figures 4.25 and 4.26 show the SHA data prediction residuals for Figures 4.23 and 4.24, indicating no systematic bias from applying the SHA to the EOF data reconstructions.

The trends discussed in this subsection define the change in the spatial distribution of the annual-oscillation throughout the solar cycle. The EOF method appears adept at isolating this signal and describing its change year-on-year. In the following subsection, I analyse the relative dominance of the annual signal for an EOF analysis of the full length of the solar cycle.

4.8.4 Long-period spatial signals throughout the solar cycle

The unavoidable inclusion of ionospheric signal in the EOF analysis is the reason that I have approached the identification of solar-cycle-related trends via isolated year-long analyses so far. EOFs temporally average spatial structure, and the strong effect of the solar cycle on the dayside (ionospheric) spatial signal is clear in the previous section. Were I to present an EOF analysis of the length of the solar cycle, I would not be able to use its temporal oscillation to describe the evolution of nightside LT-asymmetry in the long-period spatial pattern, since any solar-cycle-related modulation could simply be describing the change in the dayside signal. However, I have shown that the spatial signal of the annual oscillation is steady enough across the solar cycle for a 13-year-long EOF analysis to be of use.

Figure 4.27 shows the θ-component eigenspectrum (for the UT of 18 only) for an EOF analysis of the period 1997.0-2010.0. Note the improved signal-to-noise ratio, meaning that mode 2 is distinct enough from the others to be analysed of its own accord. Despite this, mode 1 actually explains less of the full variance than in the single-year analysis covering 1997.0-1998.0 (Figure 4.17). The first four θ-component PCs of the 13-year long analysis are shown in Figure 4.28. Note that the annual signal is still dominant in the leading mode, and that this PC has an amplitude modulation in accordance to the solar cycle (as expected from the increased ionospheric contribution). The spatial signal of this annual oscillation is shown in Figure 4.29. This reconstruction is obtained by computing the reconstruction in the same manner as in the single-year analyses discussed above (*e.g.* the data prediction shown in Figure 4.20), though in the 13 year long analyses I use the period 1997.0-1998.0 of the mode 1 PC (shown in Figure 4.28) as a 'candidate oscillation' within which the (northern hemisphere) summer signal-peak can be isolated (polynomial smoothing for the candidate oscillation is not shown). As expected from the similarity of the annual signal in each isolated year shown previously, the spatial signal of the oscillation for all years shows the usual hemispherical split, and a nightside signal peak to the dusk-side of midnight, *i.e.* the annual oscillation's spatial pattern appears dominant on the scale of the solar cycle (when the RC correction is applied). The SHA data prediction residuals for the θ-component mode 1 are shown in Figure 4.30 to be

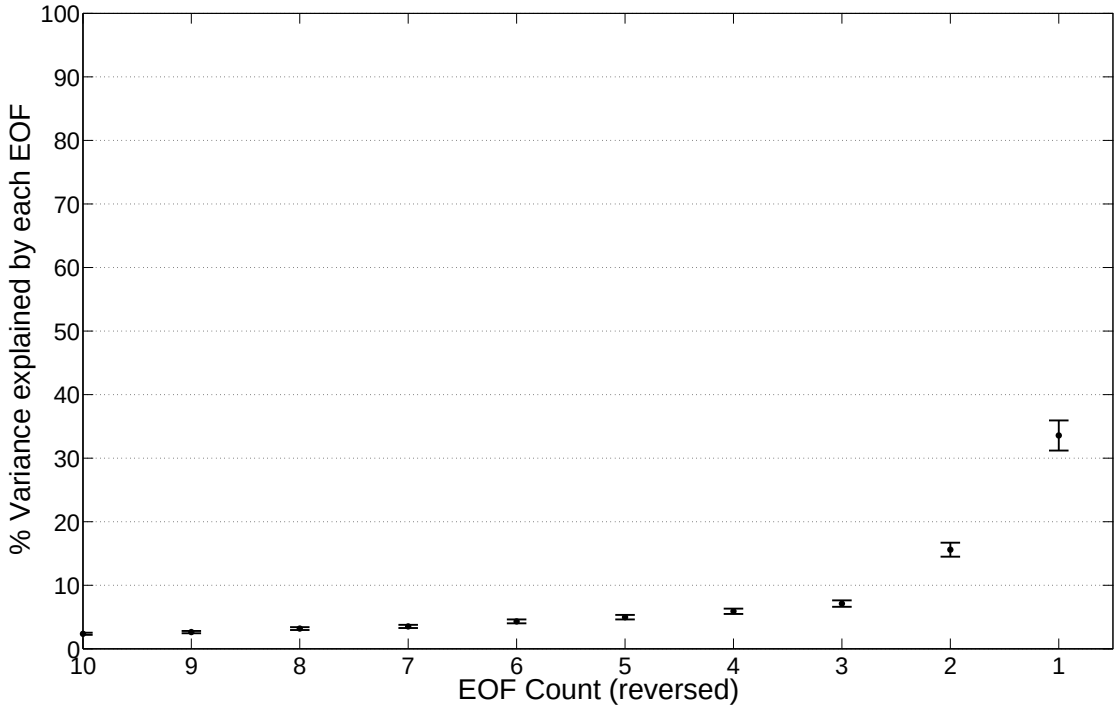

Figure 4.27: Eigenspectrum for an EOF analysis of θ-component data from 1997.0-2010.0, at a UT of 18. Only 10 of the full 142 EOF modes are shown, to better delineate the order of the leading modes.

unbiased, and do not affect the spatial pattern of the combined start-UTs.

In Figure 4.28, mode 2 still carries a (noisy) biannual signal, but is also host to a long-period trend running the length of the solar cycle. The reconstruction of this long-period oscillation is undertaken via polynomial-fitting of the full length of PC 2, as shown in Figure 4.31. The resulting spatial pattern of the mode 2 signal (for all combined start-UTs) is shown in Figure 4.32 (and the associated residuals in Figure 4.33). This spatial pattern has the same form on dayside as the synthetic example shown in Figure 4.13, indicating it to be related to the day-to-day oscillations of the ionospheric contribution (hence the biannual oscillation in this mode). This signal is expected to modulate with the solar cycle, and the combination of biannual and solar-cycle-length wavelengths in this mode is in-line with my physical interpretation of it. Note also the additional tri-lobed peaks around 22MLT in Figure 4.32. The MLT-amplitude pattern of mode 2 is quite similar to mode 1 (albeit weaker). The physical source of the mode 2 night-side peaks could either be ionospheric or magnetospheric – we cannot tell.

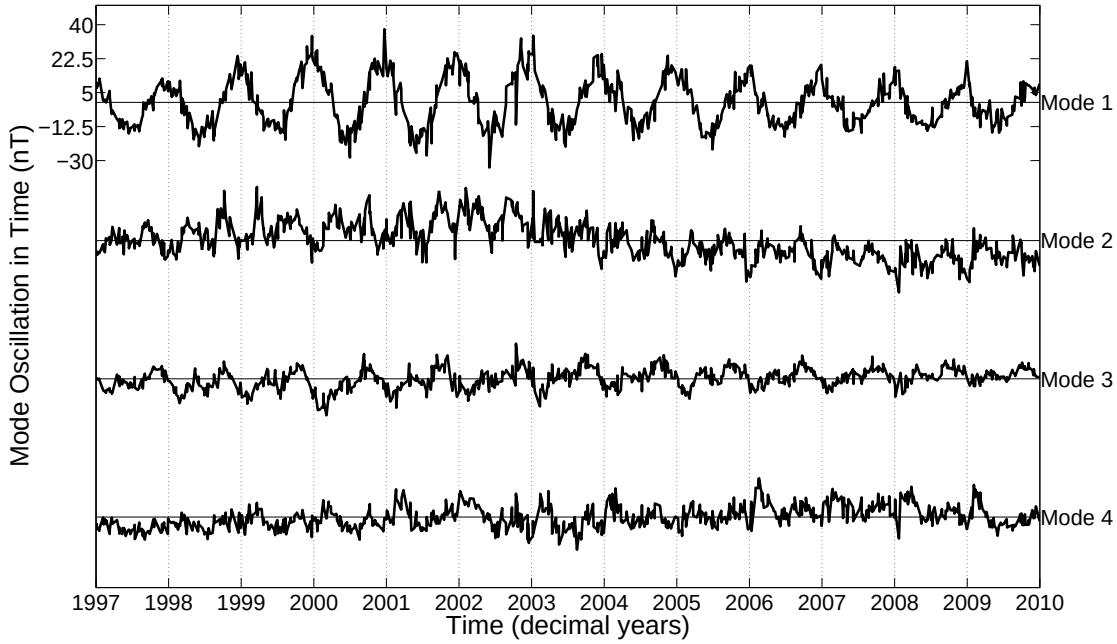

Figure 4.28: First four PCs for an EOF analysis of θ-component data from 1997.0-2010.0, at a UT of 18.

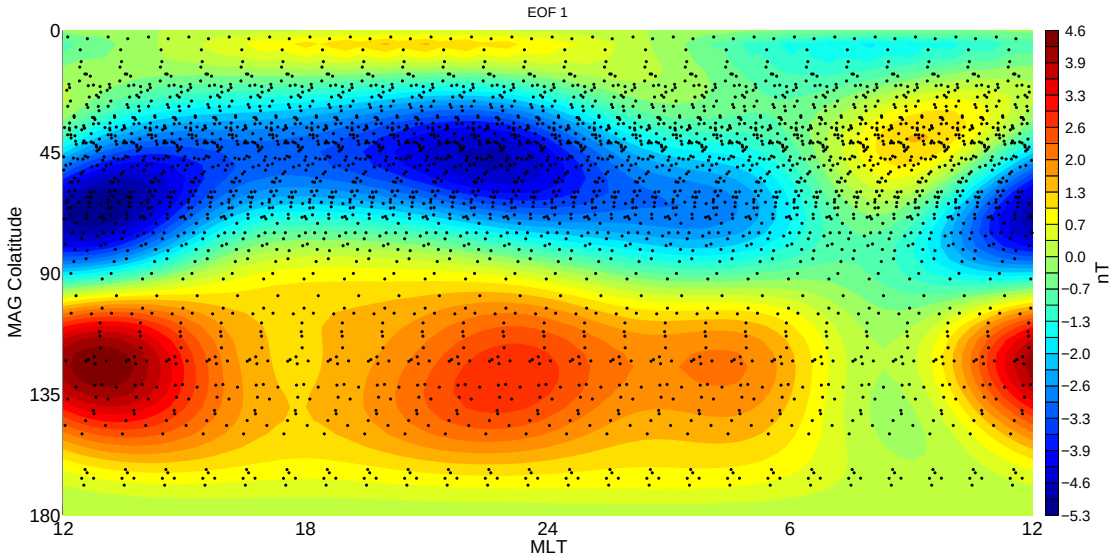

Figure 4.29: Data reconstructions of 24 θ-component EOF analyses for different start-UTs in the period 1997.0-2010.0, combined via SH analysis at degree and order 5. Black points are the distribution of input data for the SHA, derived from the 24-times replicated station distribution in LT (one distribution in LT for each of the 24 different start-UTs). Grid cell size is 5 degrees.

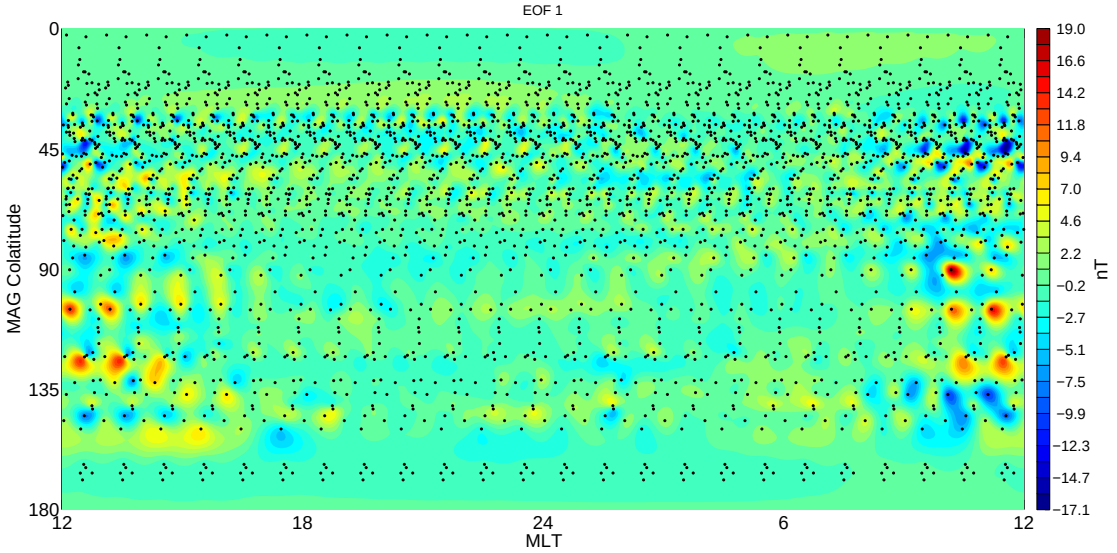

Figure 4.30: SHA data prediction residual from SHA of data reconstructions of 24 θ-component EOF analyses for different start-UTs in the period 1997.0-2010.0. Black points are the distribution of input data for the SHA, derived from the 24-times replicated station distribution in LT (one distribution in LT for each of the 24 different start-UTs). Grid cell size is 1 degree.

The results in this subsection show that the annual signal still dominates on a 13-year timespan. An analysis using only data from peak northern hemisphere summer (by selecting only quiet days in the month of July – results not shown) produces a spatial pattern similar to that of mode 1 in this instance, but with a temporal oscillation akin to mode 2 here. From the July-only analysis and the results shown here, I can state that on periods of a year or more, the spatial patterns of the dominant inducing fields are effectively static, but that their precise form will depend on the position in the solar cycle.

Due to the improvement in the signal-to-noise ratio when using all 13 years rather than just one year, it is possible to combine all start-UTs for the ϕ-component when using all 13 years of data in an EOF analysis. The PCs (for only UT 18) of this analysis (Figure 4.34) show a similar temporal progression in modes 1 and 2 to the 13-year θ-component analysis discussed above (the θ-component PCs are shown in Figure 4.28). However, from the ϕ-component eigenspectrum, it is clear that modes 2 and 3 are degenerate, so I will restrict my analysis to

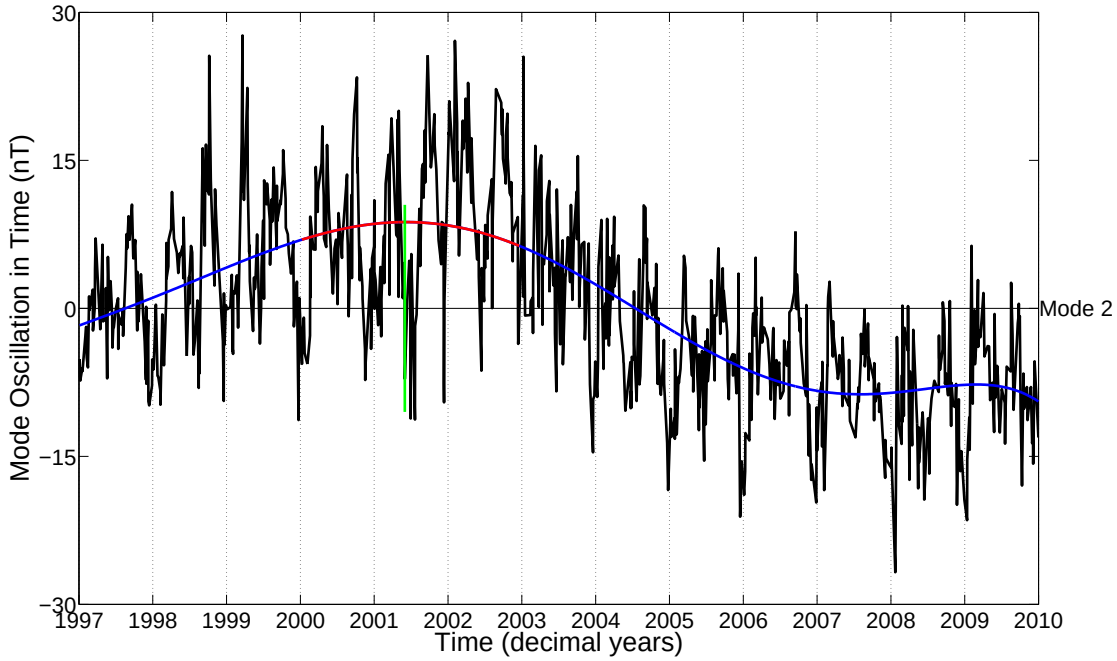

Figure 4.31: PC 2 (black line) for an EOF analysis of θ-component data from 1997.0-2010.0, at a UT of 18. The PC has been fitted with an order-6 polynomial (blue line) to reduce uncertainty in selecting the time and amplitude of the peak strength of the oscillation within the solar-cycle peak (shown in red). The time of the maximum amplitude is denoted by the green vertical line, and this controls the sign of the associated data reconstruction.

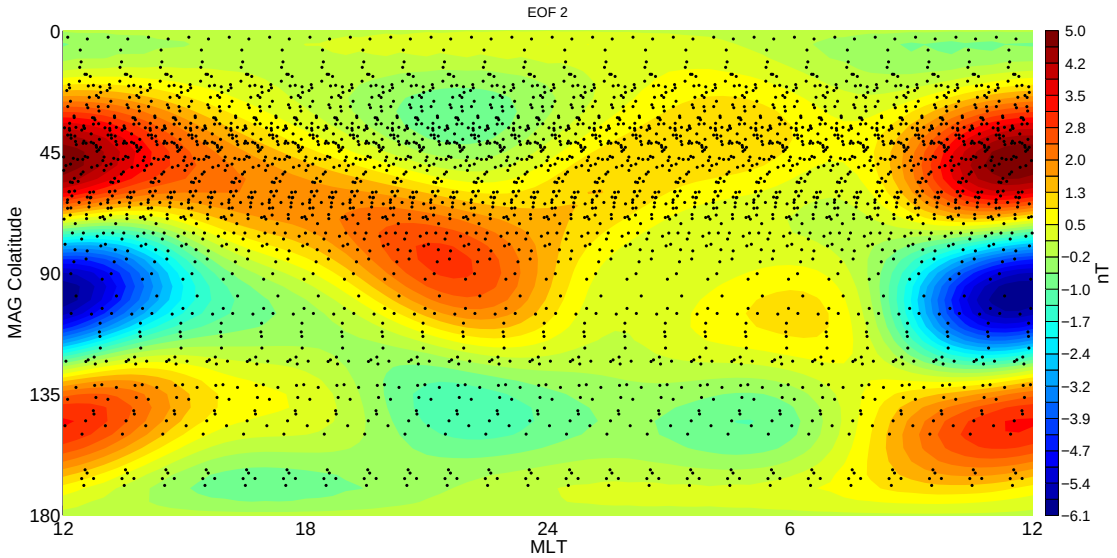

Figure 4.32: Data reconstructions of 24 θ-component EOF analyses' mode 2 for different start-UTs in the period 1997.0-2010.0, combined via SH analysis at degree and order 5. Black points are the distribution of input data for the SHA, derived from the 24-times replicated station distribution in LT (one distribution in LT for each of the 24 different start-UTs). Grid cell size is 5 degrees.

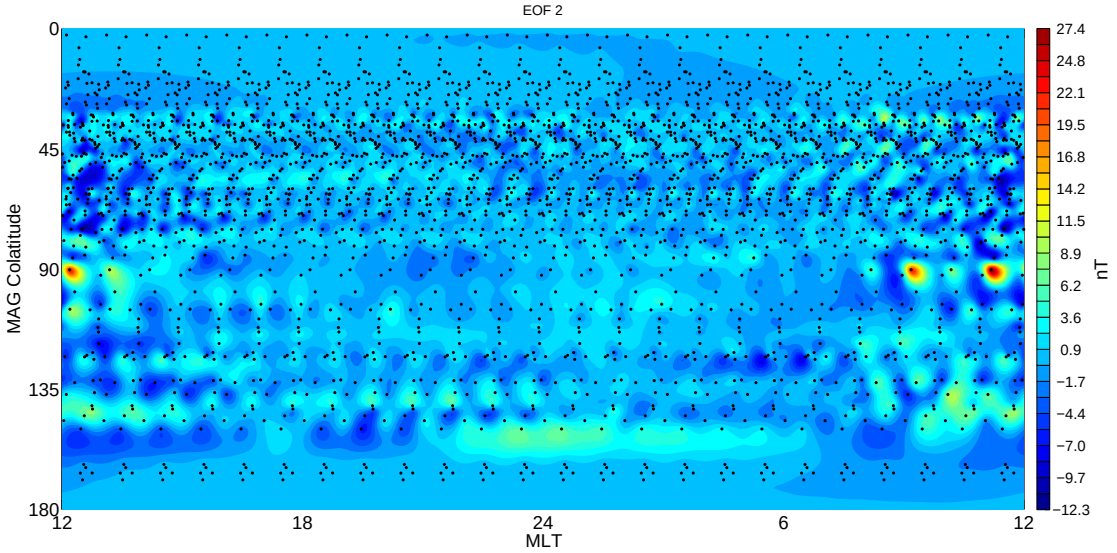

Figure 4.33: SHA data prediction residual from SHA of data reconstructions of 24 θ-component EOF analyses' mode 2 for different start-UTs in the period 1997.0-2010.0. Black points are the distribution of input data for the SHA, derived from the 24-times replicated station distribution in LT (one distribution in LT for each of the 24 different start-UTs). Grid cell size is 1 degree.

mode 1 of the 13-year ϕ-component EOF analysis here.

Figure 4.36 shows the ϕ-component data prediction from a degree and order 5 SHA of the data matrix reconstructions for all 24 EOF analyses of different start-UT, exactly as was performed for the θ-component (discussed above). The spatial pattern of mode 1 is very similar to that seen in the single-year analysis of synthetic ϕ-component data (section 4.7). The signal on the nightside is much less pronounced than it is for the equivalent θ-component analysis (Figure 4.29), though a peak in the magnitude of the pattern is apparent near dusk, indicating the same MLT dependency seen in the first two modes of the θ-component EOFs. The residuals to the ϕ-component SHA are shown to be unbiased in Figure 4.37.

The ϕ-component is dominated by dayside processes and exhibits a lesser solar-cycle-related modulation than the θ-component, so is not suitable for any further analysis related to establishing the behaviour of the LT-asymmetric fields across the solar cycle, and the θ-component remains the focus of this study.

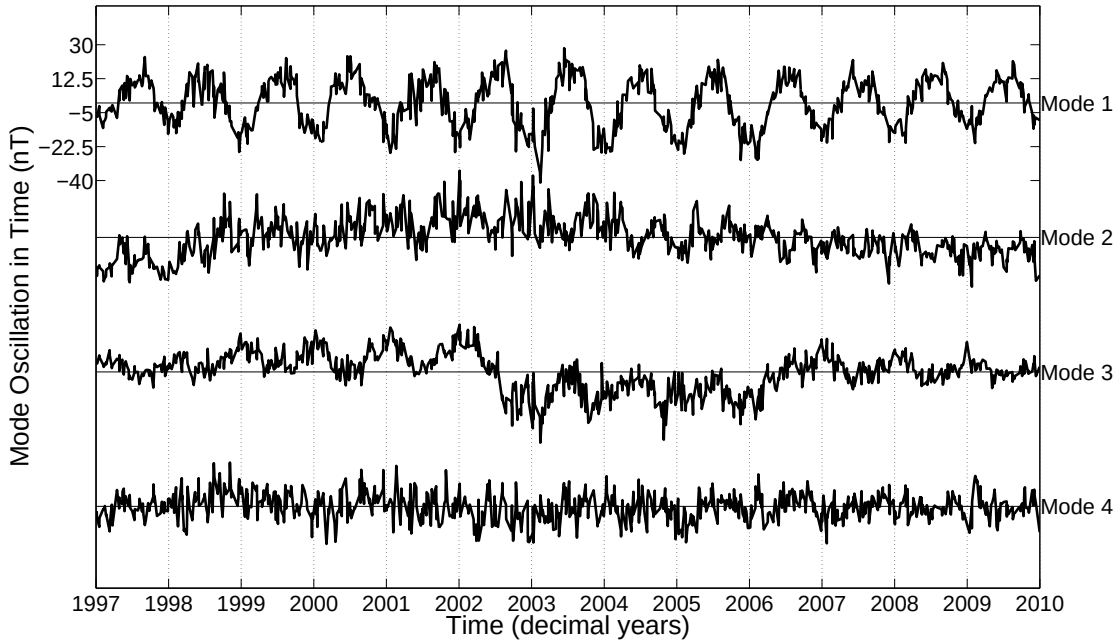

Figure 4.34: First four PCs for an EOF analysis of ϕ-component data from 1997.0-2010.0, at a UT of 18.

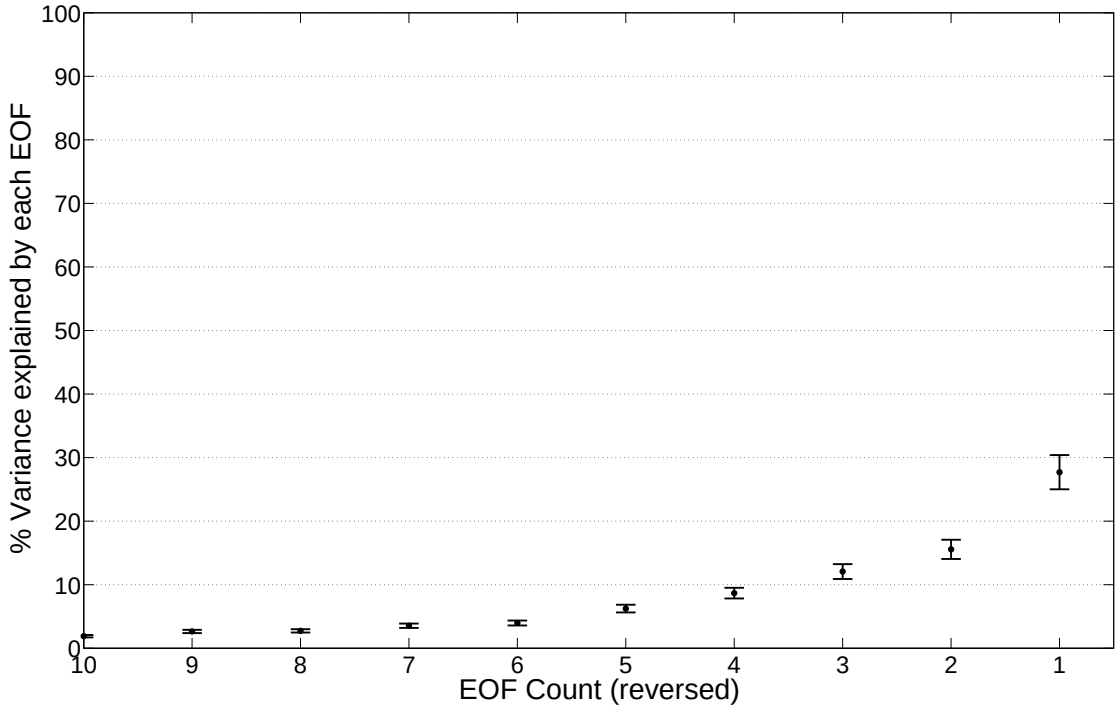

Figure 4.35: Eigenspectrum for an EOF analysis of ϕ-component data from 1997.0-2010.0, at a UT of 18. Only 10 of the full 142 EOF modes are shown, to better delineate the order of the leading modes.

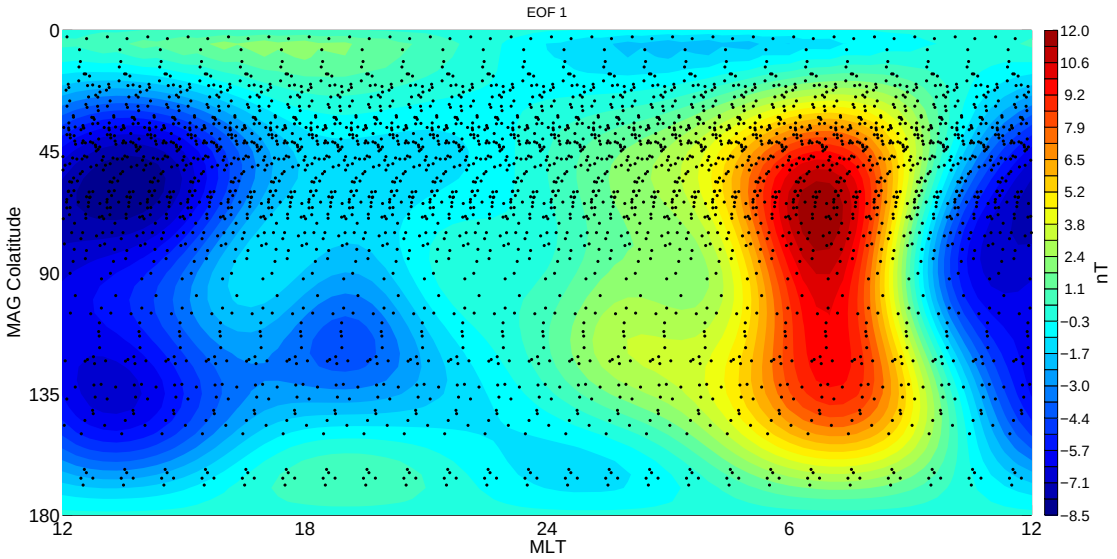

Figure 4.36: Data reconstructions of 24 ϕ-component EOF analyses for different start-UTs in the period 1997.0-2010.0, combined via SH analysis at degree and order 5. Black points are the distribution of input data for the SHA, derived from the 24-times replicated station distribution in LT (one distribution in LT for each of the 24 different start-UTs). Grid cell size is 5 degrees.

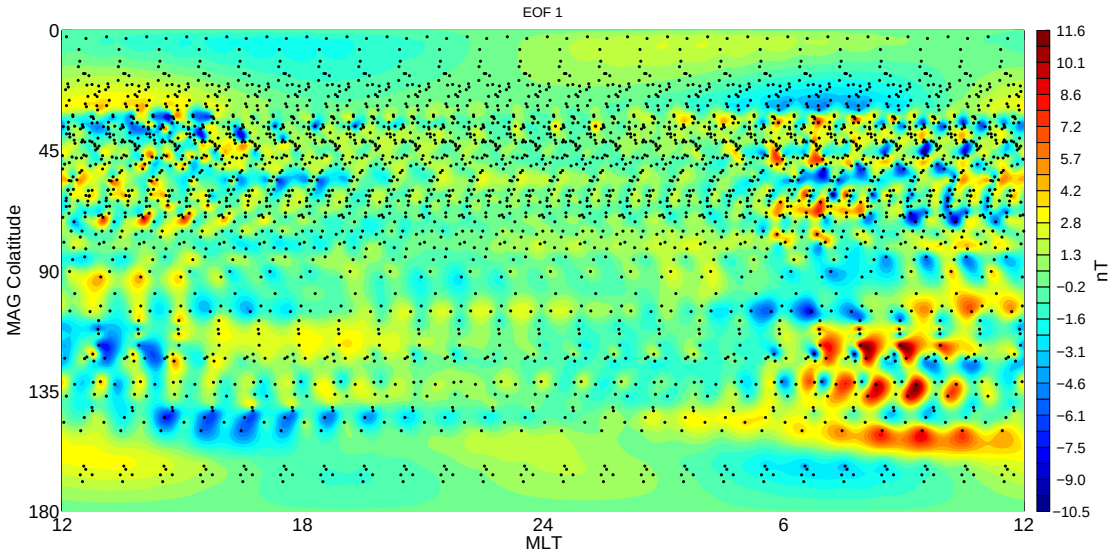

Figure 4.37: SHA data prediction residual from SHA of data reconstructions of 24 ϕ-component EOF analyses for different start-UTs in the period 1997.0-2010.0. Black points are the distribution of input data for the SHA, derived from the 24-times replicated station distribution in LT (one distribution in LT for each of the 24 different start-UTs). Grid cell size is 1 degree.

4.8.5 Isolation of the LT-symmetric external field signal

This chapter has so far focused on the specific ways of selecting and correcting the magnetic data so that the leading EOFs project onto the features I am studying, which are solely the departures from longitudinal symmetry in the long-period Earth-external magnetic fields (as stated in section 4.2). In this section, I modify the starting set-up of the EOF analysis to include the symmetric ring current signal (rather than correcting for it with the RC index as I have done previously). Whilst subsidiary to the aims of this study, this section provides useful information about the ability of the EOF analysis to isolate the signal from different oscillations of the geomagnetic field.

The eigenspectrum for an EOF analysis of θ-component data from 1997.0-2010.0 is shown in Figure 4.38. Note the improved description of the data in comparison to the percentages of variance explained in Figure 4.27. The first four PCs are shown in Figure 4.39. The temporal evolutions of modes 2 and 3 are descriptions of the same magnetic field oscillations as modes 1 and 2 in Figure 4.28, in which the RC signal was removed. The spatial signal (for all combined start-UTs) of the oscillation of mode 1 is shown in Figure 4.40, and is highly zonally symmetric on nightside, but weaker on dayside. This LT pattern is commensurate with the comprehensive analysis of the symmetric ring current undertaken by Le et al. (2004). I note that the weaker dayside intensity could also be due to the dayside magnetopause currents, the magnetic fields of which oppose the ring current field at the Earth's surface (Campbell, 2003; Maus & Lühr, 2005). The associated data prediction residuals are shown in Figure 4.41. The data reconstructions for mode 1 were performed in the same manner as in Figure 4.31, with the peak amplitude selected between 2003-2004.

Figure 4.42 shows a comparison of the mode 1 PC (red line) and the Dst index (blue line). The Pearson's correlation (*e.g.* Press, 1992) coefficient between these two data series is 0.85. In Figure 4.43, I show the (normalised) residual of the two time series in Figure 4.42 as a black line. The red line is the $F_{10.7}$ index, the blue line is the Kp index (nT equivalent) – both are normalised. Note that the parts of the mode 1 temporal evolution which are not described by

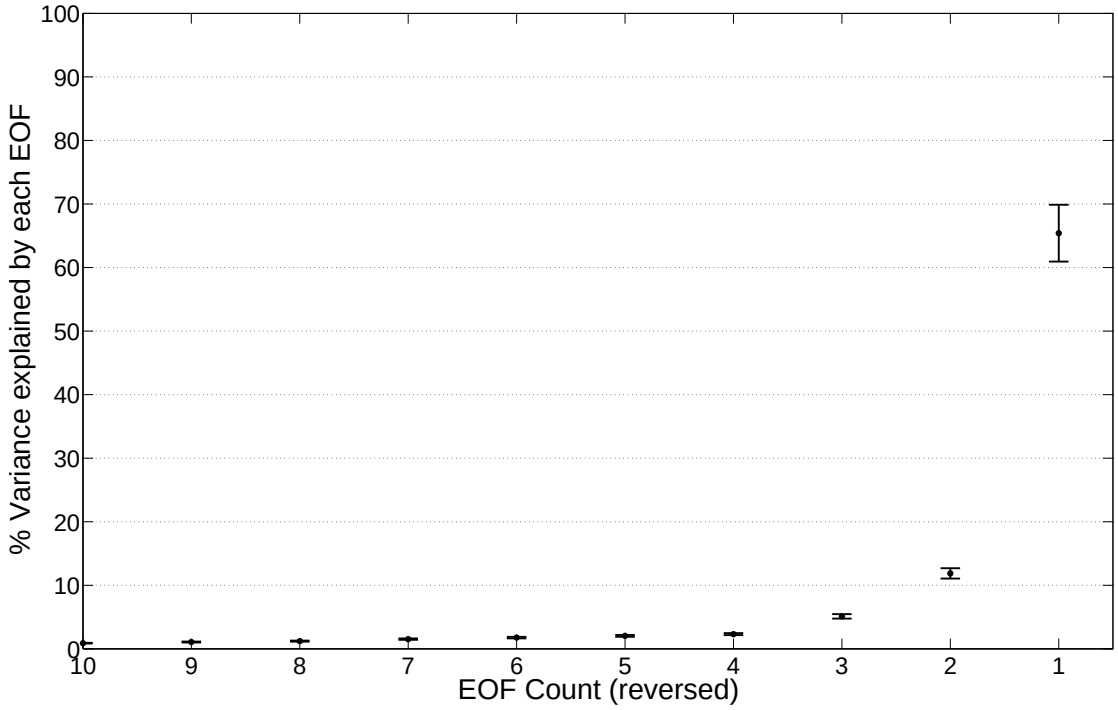

Figure 4.38: Eigenspectrum for an EOF analysis of θ-component data from 1997.0-2010.0, at a UT of 18. The RC signal has not been removed from the input data. Only 10 of the full 142 EOF modes are shown, to better delineate the order of the leading modes.

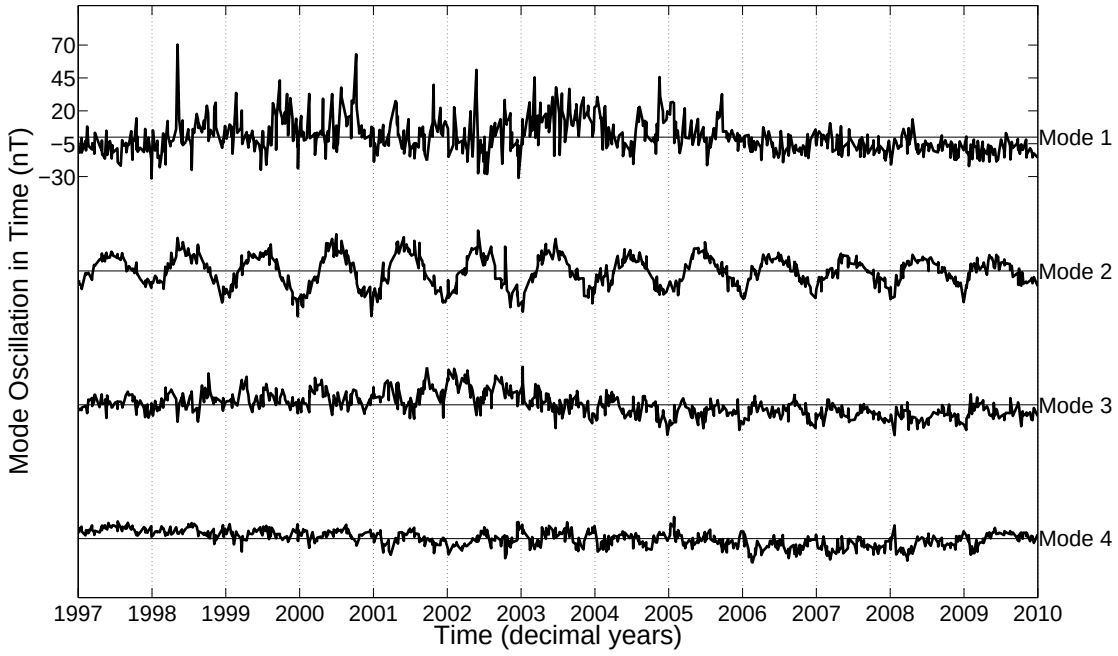

Figure 4.39: First four PCs for an EOF analysis of θ-component data from 1997.0-2010.0, at a UT of 18. The RC signal has not been removed from the input data.

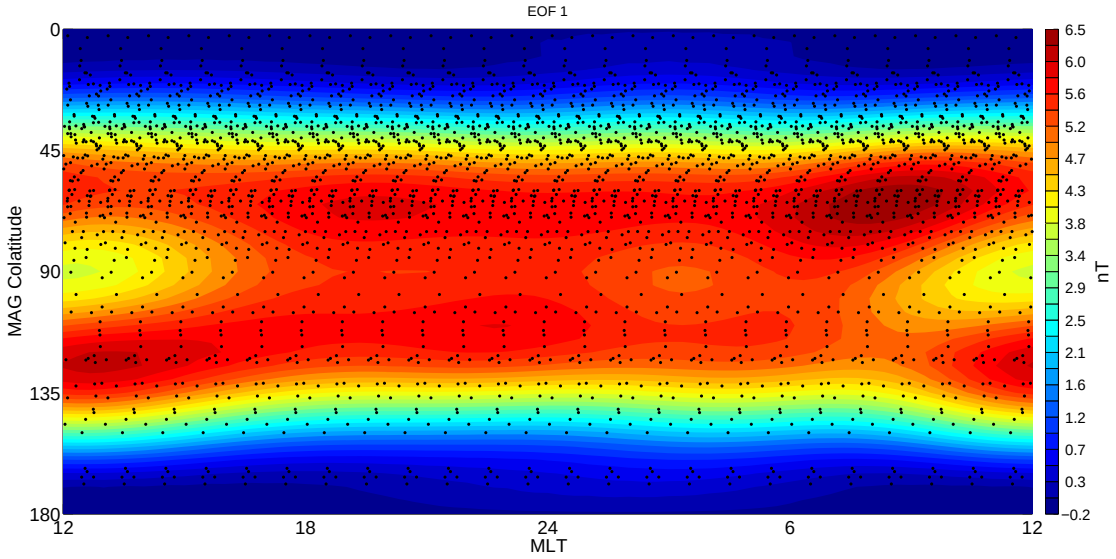

Figure 4.40: Data reconstructions of 24 θ-component EOF analyses (with the RC signal retained) for different start-UTs in the period 1997.0-2010.0, combined via SH analysis at degree and order 5. Black points are the distribution of input data for the SHA, derived from the 24-times replicated station distribution in LT (one distribution in LT for each of the 24 different start-UTs). Grid cell size is 5 degrees.

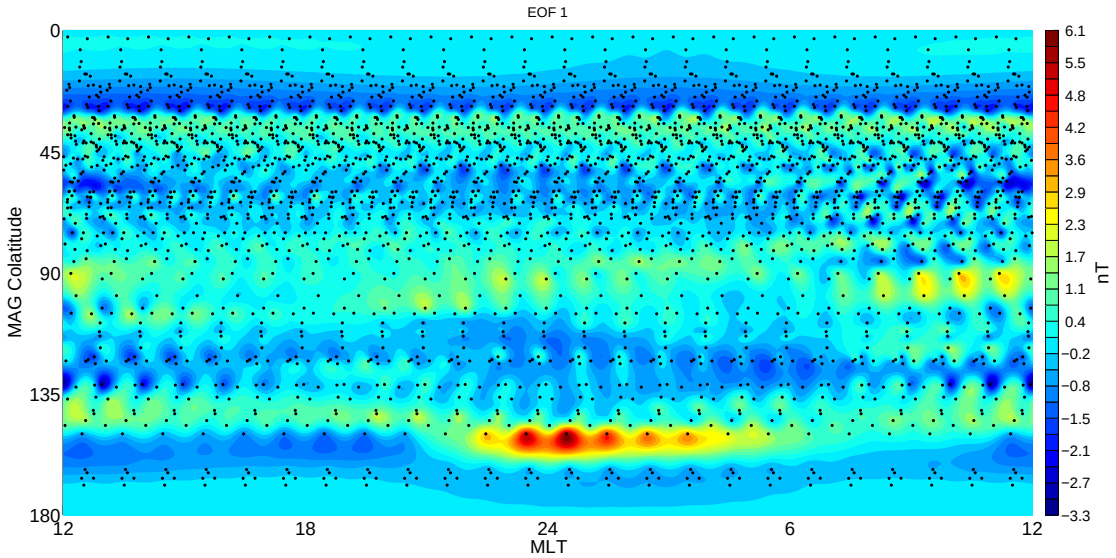

Figure 4.41: SHA data prediction residual from SHA of data reconstructions of 24 θ-component EOF analyses (with the RC signal retained) for different start-UTs in the period 1997.0-2010.0. Black points are the distribution of input data for the SHA, derived from the 24-times replicated station distribution in LT (one distribution in LT for each of the 24 different start-UTs). Grid cell size is 1 degree.

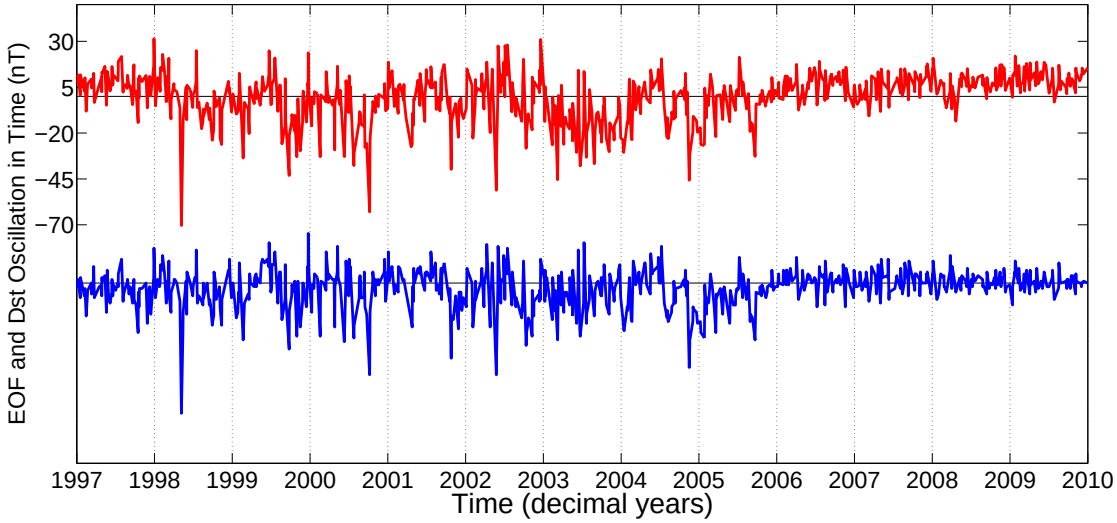

Figure 4.42: Red line is the first PC for an EOF analysis of θ-component data from 1997.0-2010.0, at a UT of 18. The RC signal has not been removed from the input data. Blue line is the Dst index. The Pearson's correlation (*e.g.* Press, 1992) value of the two time series is 0.85.

Dst appear to be a combination of the $F_{10.7}$ index (in particular the period 1997-2003), and the Kp index (specifically, the peaks in 2003-2004 and early 2008). The discontinuity in the normalised residual at 2003.0 is more clearly linked to the Kp index if this index is slightly smoothed (not shown).

As discussed by Monahan et al. (2009), EOFs are not typically expected to project uniquely onto a particular physical aspect of a system. It is instructive that regardless of the field corrections applied to the θ component, the EOF basis picks out the same patterns in the input data, albeit ones which are composed of a mixture of physical sources. Additionally, the dominance of the leading mode by a very strong, short-period signal does not appear (in this case) to affect how many modes can be trusted, nor does it affect the isolation of the long-period signals. As well as highlighting these useful aspects of how EOFs partition the magnetic field variations, the results presented in this subsection show the mixture of signals affecting the LT-symmetric field on the nightside.

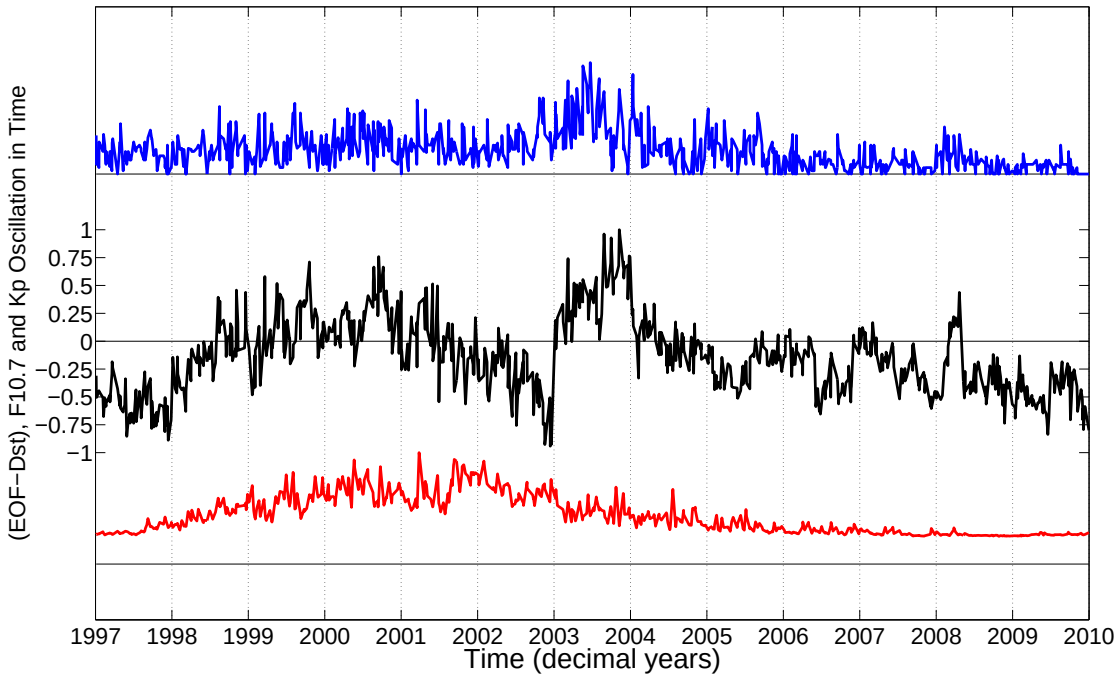

Figure 4.43: Black line is the residual of the two time series in Figure 4.42 (subsequently normalised). The red line is the normalised $F_{10.7}$ index, the blue line is the normalised Kp index (nT equivalent).

4.9 Discussion

I have shown that S-mode EOFs can provide a physically-representable decomposition of the data on an orthogonal basis, structured around the information content of the original set of time series. I have isolated the LT-asymmetric magnetic fields in the input data, and applied the EOF method to determine the dominant spatial structure in temporal progression of these data. The results have revealed the spatial pattern of the dominant long-period magnetic field oscillations, as well as how these spatial patterns change in their MLT-distribution and magnitude throughout the solar cycle. Further analyses have shown how the EOF method responds to different data-correction approaches, and what external field signals affect the LT-symmetric magnetic fields on the nightside.

The EOFs are not intended as an SHA-replacement, but have been used here to provide an alternative to the application of an SHA, providing a better representation of the field (or a more natural one) than a rigid-coordinate-system SHA could. It is possible to interpret the individual

EOFs as a physically-meaningful partition of the underlying structure of the geomagnetic field (particularly given the ability to 'reconstruct' the data for just one mode). However, the decomposition is simply a non-physical partitioning of variance – the modes are data modes, not necessarily physical modes (Bjornsson & Venegas, 1997, page 5). We should not, therefore, automatically expect the EOF modes to project uniquely onto the structures we are interested in. In the sections above I have focused only on EOF modes thought to be physically meaningful, and have assessed their significance in light of the tendency of the modes to combine physically distinct data oscillations. In this section, I assess the resolution of the decadal-period oscillations, summarise some further drawbacks of the EOF method, and describe my approach to them.

4.9.1 Isolation of decadal-period signals

As shown by the data predictions of SHAs of all combined start-UTs of separate year-long θ-component EOF analyses in Figures 4.21, 4.23 and 4.24, I have resolved the spatial signal of the annual periodicity of the external magnetic field. I have also shown, with the data predictions of an SHA of all combined start-UTs of a 13 year long EOF analysis (Figure 4.29) that a similar spatial pattern dominates the annual periodicity when its is analysed over the length of a solar cycle. The data prediction of mode 2 of the same EOF analysis (Figure 4.32) shows that the day-to-day perturbation of the fields causing the annual periodicity pattern modulate biannually, and also with a period of the length of the solar cycle.

As discussed in section 4.8.5, including the symmetric ring current signal in the EOF analysis changes the dominant mode, but does not affect the θ-component annual- and biannual-periodicity modes, nor their spatial patterns (which, although not shown in section 4.8.5, are effectively identical to the spatial patterns shown in section 4.8.4 and described in the above paragraph). The signal subspace of the EOF analysis (when including the ring current signal) spans modes 1-3, with modes 4+ exhibiting insignificant contributions to the eigenspectrum (Figure 4.38) and mathematical degeneracy, thus determined to be the noise-subspace. Further justification for the rejection of modes 4+ is given in section 4.7 – although in that case the

RC signal is removed and the noise subspace is spanned by modes 3+, the same justification applies.

The following discussion relates to the 13 year long θ-component EOF analysis in which the ring current signal was retained in the input data. An analysis of the residual of Dst and PC 1 (Figure 4.43) reveals a significant solar cycle-related modulation of the first mode. Though the amplitude of the residual is shown normalised, its true amplitude is ±20 nT. When the high-frequency 'spikes' are excluded (visually), the amplitude of the decadal-scale modulation is approximately 15 nT. Analysis of the PC of modes 2 and 3 in Figure 4.39 shows that the annual periodicity of mode 2 modulates with the solar cycle, and that mode 3's (noisy) biannual signal also has a long-period oscillation of the length of the solar cycle (this is clearer in the polynomial-fitted version of this mode, shown in Figure 4.31 where the biannual oscillation is mode 2). Visual inspection of the PC of mode 2 (annual oscillation) shows that the amplitude of its decadal-scale modulation is approximately 20 nT – this is the difference between the amplitude of the annual oscillation at solar low and at solar max. From the smoothed PC of the mode 3 biannual oscillation (Figure 4.31), the amplitude of the decadal-scale modulation of the biannual oscillation appears to be roughly 18 nT. Therefore when retaining the RC signal, the decadal-period signal spans the signal subspace with approximately equal magnitude in each of the three significant modes.

To resolve the full spatial pattern of the decadal-scale periodicity of the external magnetic fields, some combination of the patterns of the three signal subspace modes would have to be attempted. However, each EOF is sign-degenerate, and mode 2 (the annual oscillation) exhibits roughly equal magnitudes either side of zero, so combining the data matrix reconstructions of these modes into a single coherent signal would not be straightforward, nor would this add constructively to the objectives of the study. To explain why, I will consider the annual and biannual θ-component modes. From the synthetic data tests, I have shown that the biannual and annual periodicities stem from similar current sources. In the real data, a mixture of magnetospheric and ionospheric sources will contribute to these two modes. The biannual and

annual modes have also been shown (in section 4.8.4) to have similar MLT asymmetry. A combination of these two modes would not strengthen conclusions of the LT-asymmetry in the long period signal. For this reason, the long-period signal is kept as isolated modes, in which the annual periodicity is the dominant long-period oscillation. An alternative approach to resolving the decadal-period signal might be to filter out annual and sub-annual periods in the data and repeat the EOF analysis, but that would still not guarantee that the decadal-length signal dominated a single mode, so I avoid this approach.

4.9.2 Limitations of EOFs, and using them to my advantage

The EOF method, and the manner in which I have approached it, confers several limitations on my aim of isolating the long-period oscillations. Here I briefly summarise these limitations, and state how I have addressed each one, or turned it to my advantage.

Mantle induction studies typically focus solely on the symmetric magnetospheric fields (or parameterisations thereof) as the inducing source because of its spatial simplicity and dominant magnitude in the external fields. Due to the simplicity of available ionospheric models, the inclusion of the ionospheric signal is unavoidable, but identifying the signal source is not a requisite in addressing my objectives of resolving the spatial pattern of the long-period LT-asymmetric signal. The dusk intensifications could be due to ionospheric or partial ring current processes. If I was able to trust the results from individual start-UTs I could use the MAG-GEO equator difference to infer the proximity of the sources, but the station distribution is too irregular to allow this, even after the application of spatial weighting. Given the range of frequencies in the (predominantly) annual oscillation, a combined ionospheric and partial ring current source for the long period signal seems likely, in which case separating these signals is not a concern. The inclusion of the ionospheric signal is also partly responsible for the dominance of the seasonal-oscillation mode (when the symmetric ring current signal is removed). However, this is beneficial to my analysis as the north-south seasonal oscillation of the ionospheric and magnetospheric current sources places a clear zonal band (P_2^0 term) in the first mode (*e.g.* Figure 4.19), which is an ideal base upon which to look for zonal asymmetry. The

tendency of mode 1's spatial pattern (RC-removed) to be zero at the equator may not be a true reflection of the full long-period spatial signal. However, I have identified mode 2 (biannual oscillation) as a complementary part of the full pattern, and note that both modes 1 and 2 have the same asymmetry distribution in MLT, as well as a similar modulation in accordance with the solar cycle.

I am only able to use all the data via post-EOF-analysis SHA. This does bring in new challenges, but mitigates station distribution errors, avoids travelling waves and nullifies the MAG-GEO equator offset effects. Note that whilst I focus mainly on the nightside signals, I must process all MLTs in order to run the SHA (combining all start-UTs) without edge-effects near dusk. The focus on the nightside is mainly adopted to avoid the 'bleaching' (distorting) effect of the ionospheric signal maxima on the dayside. However, this approach also confers benefits to the study of the dusk sector, which is crucial to this study but would be beset by edge-effects if the data coverage stopped at 18MLT. Therefore, the requirement of a global (in the MLT frame) coverage of data by the SHA is not detrimental to my analysis. On balance, the combination of all start-UTs after their separate EOF analyses seems an effective approach.

The data distribution on the timescale of a month is controlled by the distribution of the 5 internationally-defined quietest days in each month. Therefore, I am unable to resolve temporal oscillations on the order of a month. However, the repeated resolution of the biannual signal in mode 2 is testament to the robustness of my analysis on scales on a few months or more (which is also the target periodicity). The use of polynomials in reconstructing the data matrices adds further stability (as shown in section 4.8.2).

The domain shape can have a significant impact on the patterns the EOFs project onto. Buell (1979) defined a predictable series of EOF projections which can result from the fact that EOFs are orthogonal, yet maximise variance over their spatial domain. These were later termed 'Buell patterns' (Wilks, 2005, page 480). All EOFs are spatially orthogonal to the other EOFs in the same analysis. If the leading mode is a high-magnitude signal which is nearly spatially

uniform across the area of analysis, a strong constraint is placed on the possible patterns of the non-leading modes. Buell patterns are said have arisen if the spatial patterns of the non-leading modes can be determined based on the dominance and geometry of the leading mode. Attempting to avoid Buell patterns is the reason I typically correct for the RC index signal (strong, and zonally uniform in MLT), and why the MAG poles have been removed with such a wide mask region. However, the inclusion of the RC-signal in section 4.8.5 does not appear to significantly affect the non-leading modes' projections, so I conclude that Buell patterns are not an issue in this analysis.

The static nature of the EOF basis is a potential cause for concern given the disadvantage of the invariant basis used in standard SHA methods applied to magnetic data, as noted earlier. However, I have shown that the same dominant pattern arises from the data in analyses of 1 year in length, and 13 years in length. The use of multiple separate EOF analyses throughout the solar cycle is a good basis for assessing the spatial change in the long-period signal during this period.

4.9.3 Rotation of EOFs

The effects of mathematical degeneracy on the EOFs was discussed comprehensively in section 4.7. A method commonly employed to 'separate' the patterns of mathematically degenerate (or near-degenerate) EOFs is so-called EOF rotation, discussed in Bjornsson & Venegas (1997); von Storch & Zwiers (2002); Jolliffe (2002), with Richman (1986) offering the most comprehensive treatment. Richman (1986) describes multiple approaches to EOF rotation, each a variation on the common theme of a linear recombination of the elements of the (possibly partially-degenerated) signal-subspace modes, in order to achieve what is termed 'simple structure'. The simple structure concept is met when fewer than the number of modes input into the rotation process are required to explain most of the variance of the original data. Whilst von Storch & Zwiers (2002, section 13.5.8) stated that rotation should not be applied as default, they describe its merits in reducing the orthogonality constraint on patterns contained by degenerate EOFs (since these can be arbitrarily rotated within the space that they span). I ex-

perimented with rotation (using the varimax rotation method of Kaiser (1958)) in an attempt to obtain more useful signal in cases where modes 2 and 3 were degenerate. However, the results (not shown) did not improve the usefulness of the basis expansion, and I did not use degenerate EOFs in this study.

4.10 Conclusions

The initial motivation for studying the magnetic field with EOFs was that the contamination affecting the VO estimates was not understood in terms of its ratio of zonally symmetric-to-asymmetric content. In other words, the contribution of the LT-asymmetric magnetic fields at satellite altitude to the VO solutions appeared to be less well understood than I had initially assumed. Additionally, the Ampère's integral study showed that the ionosphere might host magnetic fields stronger than previously thought, distributed systematically in LT. It became quickly apparent that the EOF expansion basis could not isolate the rapid-period fields important to the VO study's contamination. Additionally, it cannot uniquely specify (from the ground-level data required) which external fields were responsible for the resolved EOF trends, thus the confirmation of signal identified in the Ampère's integral study was not possible with this analysis. Despite these shortcomings, it was also identified that the EOF basis would be useful in isolating the spatial features of the long-period signal important to satellite magnetic data-based mantle induction studies. Since this is a key component of the *Swarm* scientific mission, this is a timely study.

In accordance with the objectives stated in section 4.2, a summary of the achievements of this study is as follows. The EOF basis can indeed provide a useful decomposition of the magnetic field, but should not be expected to uniquely determine the physical source of the magnetic fields. As shown in section 4.8.5, this can be used to my advantage, as it provides information on which physical sources combine to produce the LT-symmetric field on the nightside. I have shown that the EOF basis can be used throughout the solar cycle in periods of a year (in order to study the LT-sector dominance of the zonally asymmetric external fields) and on 13-year periods (in order to study the magnitude change in these fields).

I find that the annual periodicity of the external magnetic fields is dominated by a P_2^0 term with additional spatial amplitude peaks at local noon, and between local dusk and midnight. Of lower dominance is the EOF mode with a biannual periodicity, which has spatial amplitude peaks at the same LTs as the mode with annual periodicity, yet with no P_2^0 pattern. The dominant pattern on shorter periods is a P_1^0 term, thought to represent the symmetric ring current. Each of these three oscillations exhibits a temporal amplitude modulation with a period of the length of the solar cycle. These long-period modulations of the amplitudes of each of the three significant modes have roughly equal magnitudes, at around 15 nT.

The EOF results cannot feed directly into mantle induction studies, as they do not preserve the potential nature of the geomagnetic field. However, the isolation of the spatial signal of the long-period field should be of use to such studies. I also note from Figure 4.29 that the magnitude of the slope between the nightside peaks and dawn (*i.e.* an approximation of the magnitude of the LT-asymmetry) is roughly $1/5^{th}$ of the peak magnitude of the LT-symmetric field pattern shown in Figure 4.40. Though both patterns are weakened by the application of the SHA, I expect this magnitude-ratio between the symmetric and asymmetric fields to persist in the original data. Hence, where accurate identification of the inducing field is required, I suggest that the spatial pattern be assumed zonally asymmetric in LT. More importantly, I have shown the LT-asymmetry of the inducing field to be increasingly important on long periods, even when the LT-symmetric field is retained in the EOF analysis. The annual component of variation is spatially distinct from the symmetric ring current and has roughly equal magnitude to it. Therefore, the seasonal variation of the external field is an important factor in the long-period inducing source. Furthermore, the amplitude of the decadal-scale modulation has roughly equal magnitude to the annual signal oscillation (dependent on the point in the solar cycle). It appears from these findings that selecting the symmetric ring current as the basis for the behaviour of the inducing field on long periods ignores quite possibly the majority of the signal which actually contributes to the induction.

Chapter 5

Discussion

The results presented in this thesis have contributed to improving the state of knowledge of external magnetic fields in the near-Earth environment. Specifically, I have focused on producing advanced descriptions of the morphology of the external fields in local time, the impact of these fields on standard modelling practice, and their distribution at LEO altitudes. The choice of these foci was motivated by the challenges posed in using constellation satellite data to resolve the Earth's internal and external magnetic fields. I have sought to provide answers to questions which have surfaced throughout the course of my research (particularly from the findings of the VO study), and to address what I perceive as either long-standing problems in the discipline of geomagnetic field research, or problems which I consider will become increasingly important in the near future. Here follows a repeat of my thesis' research questions stated in chapter 1.

- To what extent are the external fields encountered in satellite magnetic data being oversimplified, and how can this be avoided in order to study the internal fields to the highest degree of accuracy?

- What is the impact of oversimplifying our description of the external fields on being able to properly model the internal field and its change in time, particularly when using spherical harmonic analysis?

- How can satellite constellation missions help improve our description of the external

magnetic fields, whilst benefiting from the advances they make possible?

In this discussion I will assess how each of the chapters 2–4 have contributed to answering these research questions, and I will set the advances I have made in the context of recent progress in the discipline of geomagnetism.

5.1 What are the recent advances in external field modelling, and what remains oversimplified?

Geomagnetic field research has benefited immensely from the continuous monitoring of the magnetic field by LEO satellites during the period 2000-2010, typically referred to as the Decade of Geopotential Field Research (DGR) (*e.g.* Friis-Christensen et al., 2009). The coverage of magnetic vector data in this decade has spanned nearly all geographic locations at the full range of local times, as well as a representative range of external field activities, in terms of the solar cycle, magnetospheric intensifications and diurnal ionospheric fluctuations. These data have led to huge improvements in the modelling of both internal and external fields, in both theory and practice. In this section I will summarise the recent advances in geomagnetic modelling with a focus on the DGR, and thereby put into broader context the advances I have made, as described in earlier chapters.

Perhaps the most innovative modelling approach developed in recent years is that of the comprehensive approach to geomagnetic field modelling. The technique was developed by Sabaka & Baldwin (1993), but has been put to its best use during the DGR, when for the first time, the Ørsted and CHAMP satellites provided temporally-sustained sampling platforms at altitudes above the Earth's surface, covering all local times. This enabled the radial separation of the ionospheric and magnetospheric magnetic field contributions in the same inversion as that in which the internal field was coestimated. The CM4 model of Sabaka et al. (2004) (the latest in the comprehensive model series) remains the most advanced separation of internal and external field sources (Olsen et al., 2010a, page 143). As discussed in chapter 1, the principal benefit of the comprehensive approach over the serial modelling approach is that the

comprehensive approach only requires a limited *a priori* knowledge of the spatial structure of the external field source regions, since this is determined in the inversion (Olsen et al., 2010a). This is true, if we assume an adequate expansion basis for the full complexity of the external field's spatial structure. However, I would argue that detailed *a priori* knowledge of the spatial structure and temporal behaviour of the external fields is essential in avoiding aliasing the contribution from these fields. The effects of external field contribution aliasing on the internal field determination by SHA were discussed in chapter 2. In short, to represent the variations of the internal and external fields simultaneously, we require a good knowledge of how the external fields (in particular) are arranged in space and time.

The parameterisation by Langel & Estes (1985) of the dominant magnetospheric field as a symmetric ring current with a quiet-time magnitude of 20 nT (near the Earth's surface) and which varies linearly with Dst, persisted until the DGR (Lühr et al., 2008b, page 49). Olsen (2002) described the method of modelling the external fields in a coordinate system close to the one in which they are naturally aligned, obtaining an improved correction for the effects of the external fields. Shortly after, a method of combining analyses of data sets which vary differently in time and space using spline curves was presented (Olsen et al., 2003b) – the same approach is used in the comprehensive models. The natural-coordinate system approach of Olsen (2002) was later adopted by Maus & Lühr (2005), who showed the quiet-time contribution of the symmetric ring current to be much smaller than reported by Langel & Estes (1985), at around 8 nT, and spatially static in SM. The authors also found a stable 13 nT contribution from a current system which is spatially static in GSM – the magnetotail and magnetopause currents. Maus & Lühr (2005) conjectured that the stable ring current field reported by Langel & Estes (1985) was in fact a superposition of separate magnetospheric sources. In summary, the data from the DGR has led to the recognition that the full magnetospheric signal contains significant contributions from structures which vary in LT, rather than being dominated by the symmetric ring current. The findings of the VO study (chapter 2) could not confirm the importance of these known LT-dependent variations to the total aliased signal. However, the strength of the asymmetric ring current has been reported by Olsen et al. (2003b) and Le et al. (2004)

and appears to be an important component of the full ring current, and the two likely vary with a similar UT dependence. It is also likely that the partial ring current supplied a significant proportion of the near-dusk intensifications in the long-period external field signal reported in the EOF study (chapter 4). Recently, Lühr & Maus (2010) have shown that whilst the GSM-orientated field reported by Maus & Lühr (2005) exhibits no solar cycle dependency, the ring current (rather, the collection of magnetospheric fields aligned in SM) exhibits a strong solar cycle dependence, with maximum magnitude 15 nT. The temporal progression of this dependence shows a good similarity to the solar-cycle dependence resolved in my EOF analysis of the dominant magnetospheric modes. The EOF approach seems suitable for determining the spatial patterns of field variability on a range of time-scales – I discuss several extra possibilities in sectin 5.3.

Given that the magnetospheric fields are ever-present at mid and low latitudes, they have typically been subject to a philosophy of removal via a correction for an estimate of their magnetic effect. The advances made during the DGR have been instrumental in refining this approach. In contrast, the ionospheric fields were until recently considered negligible on nightside, and so, presumably easier to avoid (via data selection) than the magnetospheric fields. However, in recent years a series of discoveries have helped to convey that the ionosphere plays an important role in the near-Earth magnetic environment at nightside, in addition to the well-known dayside contributions. Research into ionospheric currents using single-satellites has discovered the magnetic signature of systematic current intensifications which flow at the same altitudes of LEO satellites. Key among these efforts is the identification of the gravity and pressure-gradient currents by Lühr et al. (2003), Maus & Lühr (2006) and Alken et al. (2011). The gravity and pressure-gradient currents are a significant part of the total current resolved in the Ampère's integral study (chapter 3). At the same time as these DGR-related advances, improvements in ionospheric theory have identified new ionospheric structures at LEO altitudes. Recent research by Klimenko & Klimenko (2011, 2012); Karpachev et al. (2012) has shown the LT and latitude patterns associated with newly-identified plasma density accumulations at upper LEO altitudes, called the F3 and G layers of the ionosphere. Now that the importance of the

ionospheric currents to magnetic field modelling efforts is recognised, more effective correction strategies for the ionospheric currents can be devised. Stolle et al. (2008) have developed a proxy for the EIA strength from the EEF and the EEJ, leading to the possibility of a near-real-time index for the effect of the currents flowing in the EIA. More recently, Vanhamaki & Amm (2011) and Amm et al. (2008) have described how to use multiple simultaneously-recording satellites to produce 3-dimensional real-time models of the current contributions, allowing for more accurate corrections for their effects. The insensitivity of the in-situ estimated ionospheric electric current density (presented in chapter 3) to solar flux density (given by the $F_{10.7}$ index) indicates that determination of magnetic quiet times may require information not provided by the commonly-used indices of solar and magnetic activity. The picture of the ionosphere is one of increasing complexity, but modelling the magnetic contributions of these currents may not prove as challenging as it first appears. Pfaff (2012) suggests that the mid-to-low latitude ionosphere is significantly affected by changes in the magnetic field environment, but that the controlling factor is ultimately the atmospheric behaviour, which varies slowly. This implies that the ionosphere and its currents may be simpler to model on a first-principles basis than if they were controlled mainly by the multiple magnetic effects which impact that system – for instance from ionosphere-magnetosphere coupling. Despite this, the effects of ionosphere-magnetosphere coupling will become increasingly important to field modelling efforts, as I discuss in the following sections.

5.2 How do the external fields affect representations of the internal field?

The time change of the internal field (including its induced component) is essential for an understanding of the mechanics of formation of the magnetic field, and for producing accurate predictions of the magnetic field at future epochs. In chapter 2 I showed that aliasing of the temporally-variant external field was damaging to the internal-assumption SHAs commonly used to describe the SV, adding to the earlier study of Beggan (2009). As discussed by (*e.g.*) Olsen et al. (2010a), the SHA expansion allows for the separation of internal and external con-

tributions to the model of the field, provided that the magnetic fields are sufficiently slowly changing. Lühr et al. (2008b) stated that external field signals which are not a stable mean contribution must be parameterised time-dependently, or excluded from the SHA. When the temporal variation of the external field is not accounted for properly, the internal/external separation will no longer function effectively – in practice this means that the aliased external field contributions will appear to provide energy at spatial frequencies which the external source regions do not physically contribute to (for instance, the internal field). This introduces an erroneous radial dependence in the description of the field, and makes it unclear to what extent apparently rapid core field variations are representative of actual outer core dynamics. The magnitude of the variation of the external fields (and hence, the magnitude of their aliased signal) often matches or exceeds the magnitude of the variation of the internal fields on the (typically annual) time-scale of repeat-measurements which are used to assess internal SV. As I discussed in the previous section, a good knowledge of the morphology of the external fields is essential in any attempt to remove or separate them from the internal field contributions. I will briefly discuss the mechanisms of how the temporal variation and the spatial morphology of the external fields are commonly misrepresented.

5.2.1 Geometry of the external fields

In a GEO-fixed analysis used in the study of the Earth's internal magnetic field, aliasing of the external field signal can arise from any misrepresentation of its spatial morphology. This is because the internal and external fields are aligned in different coordinate systems – whilst the internal fields are (on short time-scales) aligned in an effectively temporally-static coordinate system, the external fields are temporally-dependent in their geographic position (Lühr et al., 2008b). Therefore even if the external fields are fixed in magnitude, if they are sampled irregularly in time at a fixed geographic position, they will lead to an aliased signal. Irregular sampling intervals are common in internal field modelling due to the aforementioned reliance of most selection criteria on quiet field periods. The majority of the aliasing associated with this effect can be removed via a correction for the estimated effects of the external fields, or

equivalently by data selection. However, as I pointed out in the previous section, the morphology of the external fields defies simple characterisation, even at quiet times. It is likely that the mis-corrected external field will either cause an aliased signal or a systematic bias in the internal field model.

5.2.2 Time-variation of the external fields

If we were to consider the external fields as spatially static (within their geographically-moving reference frame), then a bias may still arise from the fact that they are rapidly temporally variant in magnitude. This has the effect of increasing the Nyquist frequency required to represent the external fields properly in a geographically-fixed coordinate system – as discussed in chapter 2, this was the principal source of error in the VO study. This aliasing effect can be compounded by the spatial variation of the external fields, particularly if they are not corrected for properly, for instance at active times. Although the EOF study (chapter 4) focused on long-period variations, the likelihood is that the external fields identified in that study which are asymmetric in LT will have significant short-period variations in magnitude. This supposition is bolstered by the short-period departures present in the temporal evolution of (all) the leading EOF modes (though sub-daily variations are not resolved in my application of EOFs).

5.2.3 In-situ external fields at LEO altitudes

Whilst the spatial and temporal variation of the external fields can be accounted for with a precise-enough model or data selection strategy, the effect of in-situ fields at LEO altitudes is somewhat more difficult to correct for when modelling the internal field. The typical approach to describing the internal field is via a SHA, parameterising the field according to Gauss' method. As stated by Gubbins (2010) and Olsen et al. (2010a), this approach appeals because the mathematics of the expansion reflect the origin of the magnetic field, as well as exactly reproducing its behaviour in space. However, in-situ magnetic fields disrupt this approach. It is possible to account for the effect of the toroidal magnetic field generated by in-situ FACs

flowing through the measurement region. This method, also based on SHA, is called the Mie approach (Olsen et al., 2010a). Yet the Mie representation cannot account for the effects of toroidal electric currents (identified at LEO altitudes in chapter 3), which will contribute to the internal and external poloidal magnetic field representations, biasing the analysis. The precise contribution of the toroidal currents to the poloidal magnetic field representation will vary, dependent on measurement altitude and the reference altitude of the model with respect to the toroidal currents. Removal of the contributions of the ionospheric in-situ fields is a crucial step in avoiding bias in models of the internal fields, but since intensifications of these currents are still being discovered by the ionospheric community, it seems that an effective model of their magnetic effects is still some years off.

Although the DGR has prompted many advances in external field research, the discipline is still at the stage whereby only a very small fraction of the data collected in the DGR – that recorded at very quiet times – are suitable for modelling the internal field, particularly its SV (Beggan, 2009). Avoiding unwanted external field contributions is of paramount concern in being able to make full use of the data recorded by the most recent suite of LEO satellites. Even though my focus has been on mid and low-latitudes, avoiding the more complex polar regions, I have identified multiple persistent open issues hampering a full description of the external fields, and so posing an obstruction to improving core field models. In the next section I summarise these open issues and suggest the most appropriate strategies aimed at addressing each of them.

5.3 How can the description of the external fields be improved in future, in particular with constellation satellite missions?

Despite the advances in measurements and theory stated in the previous sections, attempts to robustly model the magnetic contributions from all the external field source regions are subject to substantial uncertainties, related mostly to the mechanisms of interaction between the various source regions, even at geomagnetically quiet times. Here, I outline the open issues

and persistent problems in obtaining a full picture of the external fields.

- The partial ring current: despite clear resolution of the temporal signal of this current system by Olsen et al. (2003b) and its spatial distribution by Le et al. (2004), it is not explicitly parameterised in the magnetospheric terms of CM4 (Sabaka et al., 2004), nor in the magnetospheric model of Maus & Lühr (2005). Part of the complexity in modelling this source is that it couples to the ionosphere which leads to a combination of controlling factors, making it difficult to predict.

- The effect of high geomagnetic field activity: this heavily disrupts efforts to model not only the internal fields, but also to adequately parameterise the external fields, since their morphologies and magnitudes undergo large changes within short periods of time.

- The effect of varying solar activity on long time-scales: the EOF study resolved the effects of the solar cycle on the long-period field variations. Lühr & Maus (2010) have demonstrated the magnitude of the variation of the external fields due to the solar cycle, though which source region contributes principally to this variation is not currently clear. Although a different correction for solar cycle effects is likely required at different LTs, such a model does not yet exist.

- The magnetic fields at high latitudes: as for periods of high activity, the polar regions are characterised by frequent, rapid and strong magnetic field variations.

- The applicability of commonly-used indices: whilst Dst and Kp are known not to describe the activity level adequately above mid-latitudes, there is a growing body of evidence (Olsen et al., 2003b; Lühr & Maus, 2010) to suggest that Dst (for instance) incorporates significant biases in its representation of the ring current.

- In-situ currents at LEO altitudes: these represent a significant problem in LEO altitude satellite magnetism, even at mid-latitudes. They violate the requirement of the SHA of measurements in a source-free region (Olsen et al., 2010a).

The items in the above list have varying impacts on the efforts to separate the contributions from internal and external fields in geomagnetism, but all present significant complications

to resolving the internal field and its change in time, as well as limitations on the SH degree to which the fields can be determined. The problems and open issues stated above can be summarised as stemming from either an undersampling of the inner magnetosphere when modelling the internal fields, or from shortcomings in the methods (typically SHA) used to model the external field variations. In the following sub-sections, I discuss what I consider to be the advances required to address these open issues.

5.3.1 Using constellation satellite missions to sample the inner magnetosphere

Continued measurements of the magnetic field from space are crucial to maintaining the best standards in the series of IGRF models, due mainly to the full global coverage available only from satellites (Mandea, 2006). Preparations for the launch of the *Swarm* mission (*e.g.* Friis-Christensen et al., 2006; Olsen et al., 2006a, 2007) have shown that the use of a constellation of satellites produces markedly better results than a single satellite, in the ability to resolve in-situ FACs (Ritter & Lühr, 2006), and in resolution of small-scale lithospheric features via the East-West gradient of the magnetic field (Maus et al., 2006a). The East-West gradient suppresses external fields since they vary more north-south than east-west (Friis-Christensen et al., 2006), but crucially this cannot help to characterise the geometry of the magnetospheric and ionospheric fields. Due to the compromise required to best fulfil all the science objectives of the *Swarm* mission, the chosen constellation geometry under-samples the external fields, even at the maximum LT separation of the two orbital planes. For instance, from the decomposition of the quiet-time magnetospheric signal by Maus & Lühr (2005), it is clear that the magnetosphere cannot be described as a system which responds identically in different LT sectors to various global perturbations. Furthermore, from the study of the partial ring current by Le et al. (2004), it is evident that more than two LT sectors need to be sampled simultaneously to account for the full magnetospheric signal. In summary, to effectively characterise the geometry of the magnetospheric magnetic field source regions, continuous and simultaneous measurements spaced adequately in local time and at a representative range of altitudes in the inner magnetosphere are required (Mandea, 2006). Precisely how many sectors of LT

need to be simultaneously sampled is subject to the information available on the external fields – two sectors seems inadequate given the LT-asymmetry resolved by the EOF study, and the ionospheric structures resolved by the Ampère's integral study. The limiting constraint on the required number of simultaneous measurements is likely to be the UT variation of the external fields and how large the GEO-fixed structures are that we wish to resolve. The VO study was able to shed some light on the UT variation of the external fields, but was not able to specify the UT variation of the non-symmetric magnetospheric currents. I expect that this will become clearer in the coming years, as the number of magnetometer-bearing satellites in simultaneous orbit increases.

An antidote to the large cost and complexity involved in launching a continuous stream of missions such as *Swarm* has been envisaged by (*e.g.*) Olsen et al. (2010b), Olsen & Kotsiaros (2011) and Selva & Krejci (2012). In theory, a number of small orbiters ('cubesats') could provide vector or scalar magnetic data to act as a constraint in regions where the (higher precision) larger satellites such as *Swarm* do not provide simultaneous data coverage. Ideally, the magnetometers on all the spacecraft would have a similar design in order to minimise differences in instrument sampling error. However, I have shown (in chapter 3) with data from the Ørsted and CHAMP satellites that the errors involved in the combination of datasets from different missions are minimal if the data are properly calibrated and aligned in the same reference frame. In light of the foreseeable future of satellite magnetometry depending on constellation missions, my demonstration of the multi-spacecraft integration technique for resolving LEO-altitude electric currents is timely.

In addition to *Swarm*, there have recently been a number of launches (or planned launches) of constellation satellites which carry vector magnetometers, orbiting in either LEO or in the inner magnetosphere. The CINEMA-TRIO mission (Horbury et al., 2012) is a series of spin-stabilised vector magnetometer-bearing cubesats, launched initially in late 2012, with a fourth and final orbiter due for launch in late 2013. The aim of the mission is to provide a relatively inexpensive platform for simultaneous multi-location measurements of transient magnetospheric

events (such as geomagnetic storms), which do not require high accuracy over long time-scales. The CINEMA-TRIO magnetometers have much lower accuracy than those of *Swarm*, and the cubesats have no scalar magnetometer or star camera. The orientation of the cubesat vector data is a challenge, but could be achieved (with likely substantial residual error) using the methods I discuss for the orientation of C/NOFS data, in the next section. The altitude of the CINEMA-TRIO cubesats is similar to that of *Swarm* and a combination of both missions could be a suitable proof-of-concept for an extension of the single-component gradient method (possible with *Swarm*) to a determination of the full magnetic gradient tensor (Olsen & Kotsiaros, 2011). As stated by Olsen et al. (2010b), the geometry of a constellation of satellites required to determine the full magnetic gradient tensor (even along just a single orbital path) means that a large number of satellites is required for a simultaneous global determination of this quantity. In short, an arrangement of spacecraft designed to orbit in the same LT plane and at similar altitudes (separated by no more than a few hundred km) may be inadequate to describe the electric currents at LEO altitudes, since these stem from multiple distinct regions. As I discussed above, the picture of which altitudes and LTs ionospheric current intensifications exist at is still very much in evolution. Using *Swarm* and CINEMA-TRIO in a proof-of-concept determination of electric current density could prove instrumental in the design (*e.g.* chosen altitude and geometry) of a future magnetic gradiometry mission, even if the results of the *Swarm*-cubesat combination prove to have unacceptable error. Should the electric current density at LEO altitudes be more fully determined, the LEO altitude magnetic data could be used in conjunction with inner-magnetosphere orbiters such as Cluster and the recently-launched RBSP mission to provide further data for an extension to the comprehensive modelling approach. The complexities involved in this would be substantial, but the approach deserves an investigation.

5.3.2 Modelling approaches required to make best use of the available and projected data

In chapter 1 I stated that the time-gap between the re-entry of the CHAMP satellite and the planned launch of *Swarm* could be bridged using data from the C/NOFS satellite. Owing to

the lack of an instrument for measuring the absolute magnitude of the magnetic field, the calibration applied to C/NOFS vector magnetic data removes long-period signal from both the internal and external fields. However, the C/NOFS data are available in an uncalibrated format[1], opening the possibility of a user-defined calibration, which would be different dependent on the intended use for the data. The presence of thermally-driven instrument drift makes the use of C/NOFS data for internal field studies difficult, but the vector data can be recalibrated to allow the resolution of external fields. Since the calibration requires knowledge of the (external field) information contained in the magnetic field measurement, methods to approximate this information based on a decomposition of the data itself should be explored, as follows.

An EOF-based estimate of the dominant magnetospheric signal (as resolved in chapter 4 with ground-based observatory data) could be used to orient the C/NOFS data at all LTs, as an alternative to the currently-used modified IGRF model. In order to translate the signal obtained at ground level to satellite altitudes, the EOF mode could be used as a LT-dependent scaling factor for the magnetic field prediction of (*e.g.*) the RC index (described by Olsen et al., 2011). This approach would be suitable for studying signals which are transient in LT, such as spread-F gravitational instabilities (described by *e.g.* Stolle et al., 2006). An alternative approach to the calibration would be to use the method of Holme (2000) in order to rotate the vector data into a system defined by its noisiest direction. This would define the most variant direction of the C/NOFS vector measurements, which would presumably be the axis of the major magnetospheric signal at that LT. This approach provides a better directional assessment to the EOF-based approach, since its external field information is defined from the satellite data rather than ground-based data – this approach could feasibly be used to orientate the CINEMA-TRIO data. If the method of Holme (2000) were applied to both *Swarm* and CINEMA-TRIO (or C/NOFS) vector measurements, then one of the *Swarm* satellites could be used as a remote reference for the orientation of (*e.g.*) CINEMA-TRIO data recorded in the same epochs as *Swarm* measurements. Once the rotation from the coordinate frame of maximum variance to the frame of the *Swarm* satellite star tracker is known, then the orientation of the *Swarm* star

[1] available at `http://cdaweb.gsfc.nasa.gov/istp_public/`, as of 23/05/2013

tracker can be used to orient the (*e.g.*) CINEMA-TRIO data in an ECEF frame. This assumes that the *Swarm* and CINEMA-TRIO measurements will have identical directions of maximum variance, but any difference in this in LT should be quantifiable with an EOF analysis of ground based observatory data, for the same epoch as the satellite measurements. In an orientation procedure of this type, the timescale of the analysis becomes important, as this defines which field details are removed by the rotation and so this form of vector data orientation may not be well-suited to studying transient external field signals. These calibration approaches (ground-based and in-situ LEO eigen-analyses) are both viable solutions to the problem of vector data calibration in the absence of scalar measurements, but are outside of the scope of this project.

The planned simultaneous sampling basis for magnetic data at LEO altitudes discussed in the previous sub-section should provide an excellent continuation of the measurements obtained during the DGR. In order to make best use of these data, the most appropriate modelling methods should be employed. Whilst the comprehensive SHA is still the best approach to modelling the full field (particularly the internal portion), it is perhaps not best suited for making new discoveries about the highly spatiotemporally variant external fields. Moreover, given the time (several days) taken to obtain an adequate geographic coverage from a single LEO satellite, an SHA of this data is poorly adapted to temporally parameterise field variations which occur within this time period – the external fields must either represent their time-averaged field or be corrected for with other methods (Lühr et al., 2008b). Hence SHA cannot function properly at high latitudes or periods of high activity, where strong and rapid field variations are common. In the remainder of this subsection I discuss alternatives to the SHA, and how they might be used to parameterise the external fields.

The EOF method seems well-suited to the task of defining the spatial and temporal patterns of the magnetic field in regions which defy simple characterisation. It is feasible that the EOF method (in conjunction with observatory data) could be applied to magnetic data isolated in 'bands' of different field activity levels. This would provide valuable information on the spatial distribution of the external field at active times, allowing better-targeted models of

its variations. If the levels of field activity were specified with different global indices (*e.g.* Kp, Dst), this would be useful in determining which current systems are dominant in a certain index at a given activity level. The same approach could also provide empirical information on the spatiotemporal extent and progression of global or LT-isolated disturbances such as geomagnetic storms and substorms. The EOF method would be particularly effective at high latitudes, where rapidly varying fields are very common. It would be possible to use a simple, geographically-fixed model of *Swarm* data to provide a constraint on distribution of the the auroral magnetic fields with latitude in the EOF analysis, in order to supplement the sparse observatory distribution in the polar regions. The satellite data would have poor constraint on the distribution of the magnetic fields in LT, but at these high latitudes the LT spacing of the *Swarm* satellites has a short geographic distance, mitigating this shortcoming to an extent.

The effectiveness of the EOF method on long (decadal) timescales of observatory data (as shown in chapter 4) raises the possibility of using EOFs to study mantle induction in more detail – along with Kathy Whaler I have submitted a grant proposal for this future work. Typically in studies of mantle induction, SHA is used to define a 'transfer function' relating the inducing to the induced magnetic fields, thus providing an estimate of the surrounding electrical conductivity structure (Niblett & Honkura, 1980). Whilst the inducing field is typically geometrically simple, for a 3D mantle conductivity distribution the induced field can contribute to all SH degrees and orders, leading to information loss from the truncation degree of the SHA. EOFs impose no such truncation, and could offer a much more accurate means of transfer function estimation than SHA.

Another intriguing possibility, given the planned satellite constellation missions, is the development of space-based indices for geomagnetic modelling (Lühr & Maus, 2010). The efforts of Ritter & Lühr (2006) have shown that a near real-time index of FAC structure and magnitude is possible with *Swarm*. This could provide a good indication of ionosphere-magnetosphere interaction at mid-latitudes on the nightside – I identified this as a region requiring further study in chapter 3, particularly since the dependence of ionospheric magnetic fields on the (scaled)

$F_{10.7}$ index appears less straightforward than typically assumed. An index of the FACs could also be used as a proxy to describe the strength and position of the partial ring current (which closes in the ionosphere via these FACs), though this parameterisation may not prove adequate as a direct correction for the effects of the partial ring current system. Another promising possibility is the development of a series of half-orbit averages of satellite magnetic data, in order to solve for the magnetospheric dipole component (currently being developed for *Swarm* by Brian Hamilton at the British Geological Survey). This quantity is likely variant in LT, and could adequately correct for the partial ring current effects (again, in near real-time). However, the geometry of the magnetospheric dipole is likely to change between seasons and throughout the solar cycle, even at the same local time. A series of monthly EOF analyses of ground-based observatory data could specify the geometry of the leading magnetospheric variation. From this, an empirical coordinate system could be defined in order to best-fit the length of satellite track used for the magnetospheric dipole determination, improving its accuracy and accounting for long-period changes in the external fields. In this way, ground-based and satellite data would be used together to provide a better solution for external field noise when modelling the internal field.

In addition to the plans for describing the partial ring current discussed above, it is likely that additional scaling laws and proxies can be developed for other external field current systems with the planned new data. Most desirable would be the ability to link new developments in the state of external field knowledge to established quantities (such as the Dst index), as I discussed above. It would be interesting to study the possibility of linking a certain Dst/Kp ratio to a known current system intensification which is not necessarily described well by either of the indices themselves. In this way, proxy-corrections for newly-discovered current systems could be applied in epochs in which these current systems were not explicitly measured – for instance, during the DGR. A good first-step would be the use of EOFs to better determine which current systems contribute to the well-known global indices.

When a more suitable correction for the external field effects is possible, the data from the

DGR will be able to be used in full, and modelling methods such VOs could work effectively. My research has contributed to answering the open issues stated at the top of this section, but the same core issues remain. The approaches laid out in this chapter should help define how best to address the challenge of separating internal and external magnetic field contributions in the near future.

5.4 Summary and conclusions

My research for this thesis started with an application of the Virtual Observatory (VO) method (of Mandea & Olsen (2006)) to a set of synthetic magnetic vector data prepared by Olsen et al. (2007) for the upcoming *Swarm* constellation. The aim was to assess why an earlier (Beggan et al., 2009) study had shown high-degree sectorial banding in globally-distributed local potential field solutions of real data, which adopted an internal field assumption. Using synthetic data (*i.e.* simplified from the real case), my study confirmed the existence of the sectorial banding (indicating a persistent effect). I showed the precise causes and effects of the aliasing of external fields on an internal-assumption SHA. However, the study was not able to reliably determine how much of the total aliased signal was due to the magnetic fields which are asymmetric in local time (LT), and those which are LT-symmetric but which change in Universal Time (UT). Further studies were then undertaken to review the structure of the external field in more detail.

I investigated the possibility that the (VO) sectorial banding could be due to magnetic fields generated at satellite altitude interfering with the local potential field solutions – this is also of general scientific interest, and is a timely investigation. I used data from times when Ørsted and CHAMP orbits approximately coincided, with CHAMP below Ørsted, to calculate Ampère's integral, as a method for inferring the mid-latitude ionospheric current density. The results have been compared with the output of the CTIP (Coupled Thermosphere-Ionosphere-Plasmasphere) Model, a 3-dimensional numerical model of ionospheric composition and temperature. The Ampère's integral study resolved significant zonal electric current flow in the region between the two satellites. The method revealed apparently new ionospheric features,

showing that regions thought free of currents are in fact host to systematic current intensifications. However, the distribution of these currents was equatorially-focused and they could not have contributed dominantly to the sectorial banding in the VOs, which was much wider in north-south extent and stronger in amplitude.

I assessed the LT distribution of the external fields as seen at ground altitude using the method of Empirical Orthogonal Functions (EOFs). The EOF application successfully resolved the spatio-temporal distribution of the LT-asymmetric external magnetic fields over the course of a full solar cycle. The study showed that the LT distribution of the external magnetic fields changes with solar activity, providing a possible source of aliasing if these fields are improperly corrected for. The EOF study provided further evidence that the LT-asymmetry of the external fields should be accounted for in order to correct for them accurately, and to be able to use the data from the last decade to its full potential.

The methods developed and presented in this thesis show promise for future application in geomagnetism, particularly in processing the anticipated satellite constellation data to best effect. I expect that the work I have presented here will prove useful in future studies of the external fields, and in combining satellite and ground observatory data in the study of the internal field.

Bibliography

Akasofu, S. & Kamide, Y., 1987. *The Solar Wind and the Earth*, Geophysics and Astrophysics Monographs, Springer.

Alken, P. & Maus, S., 2010. Electric fields in the equatorial ionosphere derived from CHAMP satellite magnetic field measurements, *Journal of Atmospheric and Solar-Terrestrial Physics*, **72**(4), 319–326.

Alken, P., Maus, S., Richmond, A. D., & Maute, A., 2011. The ionospheric gravity and diamagnetic current systems, *Journal of Geophysical Research-Space Physics*, **116**, 9.

Amm, O., Aruliah, A., Buchert, S. C., Fujii, R., Gjerloev, J. W., Ieda, A., Matsuo, T., Stolle, C., Vanhamaki, H., & Yoshikawa, A., 2008. Towards understanding the electrodynamics of the 3-dimensional high-latitude ionosphere: present and future, *Annales Geophysicae*, **26**(12), 3913–3932.

Amm, O., Fujii, R., Kauristie, K., Aikio, A., Yoshikawa, A., Ieda, A., & Vanhamki, H., 2011. A statistical investigation of the Cowling channel efficiency in the auroral zone, *J. Geophys. Res.*, **116**(A2), A02304.

Appleton, E. V., 1946. 2 anomalies in the ionosphere, *Nature*, **157**(3995), 691–691.

Auster, H., Mandea, M., Hemshorn, A., Pulz, E., & Korte, M., 2007. Automation of absolute measurement of the geomagnetic field, *Earth, Planets, and Space*, **59**(9), 1007–1014.

Backus, G., 1986. Poloidal and toroidal fields in geomagnetic field modeling, *Reviews of Geophysics*, **24**(1), 75–109.

Backus, G., Parker, R., & Constable, C., 1996. *Foundations of Geomagnetism*, Cambridge University Press.

Bailey, G. J., Balan, N., & Su, Y. Z., 1997. The Sheffield University plasmasphere ionosphere model - a review, *Journal of Atmospheric and Solar-Terrestrial Physics*, **59**(13), 1541–1552.

Balan, N. & Bailey, G. J., 1995. Equatorial plasma fountain and its effects - possibility of an additional layer, *Journal of Geophysical Research-Space Physics*, **100**(A11), 21421–21432.

Balan, N. & Bailey, G. J., 1995. *Modeling studies of equatorial plasma fountain and equatorial anomaly*, vol. 18 of **Advances in Space Research**, pp. 107–116, Pergamon Press Ltd, Oxford.

Balasis, G. & Egbert, G. D., 2006. Empirical orthogonal function analysis of magnetic observatory data: Further evidence for non-axisymmetric magnetospheric sources for satellite induction studies, *Geophysical Research Letters*, **33**(11), 4.

Baldwin, M. P., Stephenson, D. B., & Jolliffe, I. T., 2009. Spatial Weighting and Iterative Projection Methods for EOFs, *Journal of Climate*, **22**(2), 234–243.

Baumjohann, W. & Treumann, R., 1997. *Basic Space Plasma Physics*, Imperial College Press.

Beckers, J. M. & Rixen, M., 2003. EOF Calculations and Data Filling from Incomplete Oceanographic Datasets, *Journal of Atmospheric and Oceanic Technology*, **20**(12), 1839–1856.

Beggan, C., 2009. *Secular Variation Prediction of the Earth's Magnetic Field Using Core Surface Flows*, Ph.D. thesis, Univ. Edinburgh.

Beggan, C. D., Whaler, K. A., & Macmillan, S., 2009. Biased residuals of core flow models from satellite-derived 'virtual observatories', *Geophysical Journal International*, **177**(2), 463–475.

Bjornsson, H. & Venegas, S., 1997. A manual for EOF and SVD analyses of climatic data, *CCGCR (Centre for Climate and Global Change Research) Report*, **97**(1).

Boas, M., 2006. *Mathematical methods in the physical sciences*, vol. 2, Wiley.

Buell, C. E., 1979. Physical interpretation of Empirical Orthogonal Functions, *Bulletin of the American Meteorological Society*, **60**(5), 587–587.

Buffett, B. A., 2003. The Thermal State of Earth's Core, *Science*, **299**(5613), 1675–1677.

Campbell, W., 2003. *Introduction to geomagnetic fields*, Cambridge Univ Pr.

Cattaneo, F. & Hughes, D. W., 2001. Solar dynamo theory: a new look at the origin of small-scale magnetic fields, *Astronomy & Geophysics*, **42**(3), 3.18–3.22.

Chave, A., Thomson, D., & Ander, M., 1987. On the robust estimation of power spectra, coherences, and transfer functions, *J. geophys. Res*, **92**(B1), 633–648.

Chisham, G. & Freeman, M. P., 2010. On the non-Gaussian nature of ionospheric vorticity, *Geophysical Research Letters*, **37**(12).

Colomb, F., Alonso, C., Hofmann, C., & Nollmann, I., 2004. SAC-C mission, an example of international cooperation, *Advances in Space Research*, **34**(10), 2194 – 2199.

Constable, C. & Constable, S., 2004. Satellite magnetic field measurements: Applications in studying the deep Earth, *Geophysical Monograph*, **150**, 147159.

Covington, A., 1969. Solar radio emission at 10.7 cm, 1947-1968, *Journal of the Royal Astronomical Society of Canada*, **63**, 125.

Cowling, T., 1932. Magnetism, solar: The electrical conductivity of an ionised gas in the presence of a magnetic field, *Monthly Notices of the Royal Astronomical Society*, **93**, 90.

Crain, D. J., Heelis, R. A., & Bailey, G. J., 1993. Effects of electrical coupling on equatorial ionospheric plasma motions - when is the F-region a dominant driver in the low-latitude dynamo?, *Journal of Geophysical Research-Space Physics*, **98**(A4), 6033–6037.

Daglis, I. A., Thorne, R. M., Baumjohann, W., & Orsini, S., 1999. The terrestrial ring current: Origin, formation, and decay, *Reviews of Geophysics*, **37**(4), 407–438.

Dunham, W., 1990. *Journey through genius: the great theorems of mathematics*, Wiley science editions.

Eccles, J. V., 2004. The effect of gravity and pressure in the electrodynamics of the low-latitude ionosphere, *Journal of Geophysical Research-Space Physics*, **109**(A5), 7.

Eshel, G., 2012. *Spatiotemporal data analysis*, Princeton University Press, Princeton.

Fejer, B. G., Souza, J. R., Santos, A. S., & Pereira, A. E. C., 2005. Climatology of F region zonal plasma drifts over Jicamarca, *Journal of Geophysical Research-Space Physics*, **110**(A12), 10.

Finlay, C. C., Maus, S., Beggan, C. D., Bondar, T. N., Chambodut, A., Chernova, T. A., Chulliat, A., Golovkov, V. P., Hamilton, B., Hamoudi, M., Holme, R., Hulot, G., Kuang, W., Langlais, B., Lesur, V., Lowes, F. J., Lühr, H., Macmillan, S., Mandea, M., McLean, S., Manoj, C., Menvielle, M., Michaelis, I., Olsen, N., Rauberg, J., Rother, M., Sabaka, T. J., Tangborn, A., Tøffner-Clausen, L., Thebault, E., Thomson, A. W. P., Wardinski, I., Wei, Z., & Zvereva, T. I., 2010. International Geomagnetic Reference Field: the eleventh generation, *Geophysical Journal International*, **183**(3), 1216–1230.

Finlay, C. C., Jackson, A., Gillet, N., & Olsen, N., 2012. Core surface magnetic field evolution 2000-2010, *Geophysical Journal International*, **189**(2), 761–781.

Fleisch, D., 2008. *A Student's Guide to Maxwell's Equations*, Cambridge University Press.

Forbes, J. M., 1981. The equatorial electrojet, *Reviews of Geophysics*, **19**(3), 469–504.

Friis-Christensen, E., Lühr, H., & Hulot, G., 2006. *Swarm*: A constellation to study the Earth's magnetic field, *Earth Planets and Space*, **58**(4), 351–358.

Friis-Christensen, E., Lühr, H., Hulot, G., Haagmans, R., & Purucker, M., 2009. Geomagnetic research from space, *Eos*, **90**, 25.

Glassmeier, K.-H., Soffel, H., & Negendank, J., 2008. *Geomagnetic Field Variations*, Advances in Geophysical and Environmental Mechanics and Mathematics, Springer.

Glatzmaier, G. A. & Olson, P., 2005. Probing the geodynamo, *Scientific American*, **292**(4), 50–57.

Goldberg, R. A., 1965. Equatorial geomagnetic anomaly and its associated current system, *Journal of Geophysical Research*, **70**(21), 5417–5424.

Golovkov, V., Zvereva, T., & Chernova, T., 2007. Space-time modeling of the main magnetic field by combined methods of spherical harmonic analysis and natural orthogonal components, *Geomagnetism and Aeronomy*, **47**, 256–262.

Gubbins, D., 1983. Geomagnetic field analysis - I. Stochastic inversion, *Geophysical Journal International*, **73**(3), 641–652.

Gubbins, D., 2010. Terrestrial magnetism: Historical perspectives and future prospects, *Space Science Reviews*, **155**(1-4), 9–27.

Gubbins, D. & Herrero-Bervera, E., 2007. *Encyclopedia of geomagnetism and paleomagnetism*, Kluwer Academic Pub.

Hannachi, A., Jolliffe, I. T., & Stephenson, D. B., 2007. Empirical orthogonal functions and related techniques in atmospheric science: a review, *International Journal of Climatology*, **27**(9), 1119–1152.

Hapgood, M. A., 1992. Space Physics Coordinate Transformations - a User Guide, *Planetary and Space Science*, **40**(5), 711–717.

Hedin, A. E., 1987. MSIS-86 Thermospheric Model, *Journal of Geophysical Research: Space Physics*, **92**(A5), 4649–4662.

Heelis, R. A., 2004. Electrodynamics in the low and middle latitude ionosphere: a tutorial, *Journal of Atmospheric and Solar-Terrestrial Physics*, **66**(10), 825–838.

Heelis, R. A., Kendall, P. C., Moffett, R. J., Windle, D. W., & Rishbeth, H., 1974. Electrical coupling of E-regions and F-regions and its effect on F-region drifts and winds, *Planetary and Space Science*, **22**(5), 743–756.

Holme, R., 2000. Modelling of attitude error in vector magnetic data: application to ørsted data, *Earth Planets and Space*, **52**(12), 1187–1197.

Horbury, T. S., Brown, P., Eastwood, J. P., Archer, M., Lin, R. P., Immel, T., Glaser, D., Lee, D.-H., Seon, J., & Jin, H., 2012. CINEMA/TRIO A three-spacecraft space weather CubeSat mission, presented at NAM (National Astronomy Meeting), Manchester, UK, March 27-30.

Huang, C., 1974. Certain behavior of ionospheric F2 region at low latitudes, *Radio Science*, **9**(5), 519–532.

Jackson, A., Jonkers, A. R. T., & Walker, M. R., 2000. Four centuries of geomagnetic secular variation from historical records, *Philosophical Transactions of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences*, **358**(1768), 957–990.

Jankowski, J. & Sucksdorff, C., 1996. Guide for Magnetic Measurements and Observatory Practice, *Int. Assoc. of Geomagn. and Aeron., Boulder, Colo*.

Jolliffe, I. T., 2002. *Principal Component Analysis*, Springer, Berlin, 2nd edn.

Jones, C. A., 2011. *Planetary Magnetic Fields and Fluid Dynamos*, vol. 43 of **Annual Review of Fluid Mechanics**, pp. 583–614, Annual Reviews, Palo Alto.

Jones, C. A., Thompson, M. J., & Tobias, S. M., 2010. The Solar Dynamo, *Space Science Reviews*, **152**(1-4), 591–616.

Juusola, L., Amm, O., Kauristie, K., & Viljanen, A., 2007. A model for estimating the relation between the Hall to Pedersen conductance ratio and ground magnetic data derived from CHAMP satellite statistics, in *Annales Geophysicae*, vol. 25, pp. 721–736.

Kaiser, H., 1958. The varimax criterion for analytic rotation in factor analysis, *Psychometrika*, **23**(3), 187–200.

Karpachev, A. T., Klimenko, M. V., Klimenko, V. V., Zhbankov, G. A., & Telegin, V. A., 2012. Latitudinal structure of the equatorial F3 layer based on Intercosmos-19 topside sounding data, *Journal of Atmospheric and Solar-Terrestrial Physics*, **77**, 186–193.

Kelley, M., 2009. *The Earth's ionosphere: plasma physics and electrodynamics*, vol. 96, Academic Pr.

Kivelson, M. G. & Russell, C. T., 1995. *Introduction to space physics*, Cambridge Univ Pr.

Klimenko, M. & Klimenko, V., 2012. Mechanisms of stratification of the F2 layer and formation of the F3 and G layers in the equatorial ionosphere, *Geomagnetism and Aeronomy*, **52**, 321–334.

Klimenko, M. V. & Klimenko, V. V., 2011. Numerical Simulation of the F2-Layer Stratification and Appearance of the F3 and G Layers in the Equatorial Ionosphere: The Morphology of the Phenomena, *Geomagnetism and Aeronomy*, **51**(5), 646–655.

Kuvshinov, A. & Semenov, A., 2012. Global 3-D imaging of mantle electrical conductivity based on inversion of observatory C-responses - I. An approach and its verification, *Geophysical Journal International*, **189**(3), 1335–1352.

Langel, R. A., 1987. The main field, *Geomagnetism*, **1**(4).

Langel, R. A. & Estes, R. H., 1985. The near-Earth Magnetic-Field at 1980 determined from Magsat Data, *Journal of Geophysical Research*, **90**(NB3), 2495–2509.

Langel, R. A. & Hinze, W. J., 1998. *The magnetic field of the Earth's lithosphere: the satellite perspective*, Cambridge University Press.

Le, G., Russell, C. T., & Takahashi, K., 2004. Morphology of the ring current derived from magnetic field observations, *Annales Geophysicae*, **22**(4), 1267–1295.

Le, G., Burke, W. J., Pfaff, R. F., Freudenreich, H., Maus, S., & Lhr, H., 2011. C/NOFS measurements of magnetic perturbations in the low-latitude ionosphere during magnetic storms, *Journal of Geophysical Research: Space Physics*, **116**(A12), 11.

Leopardi, P., 2007. *Distributing points on the sphere: partitions, separation, quadrature and energy*, Ph.D. thesis, University of New South Wales.

Lowes, F. J., 1974. Spatial Power Spectrum of the Main Geomagnetic Field, and Extrapolation to the Core, *Geophysical Journal of the Royal Astronomical Society*, **36**(3), 717–730.

Lühr, H. & Maus, S., 2006. Direct observation of the F region dynamo currents and the spatial structure of the EEJ by CHAMP, *Geophysical Research Letters*, **33**(24), 6.

Lühr, H. & Maus, S., 2010. Solar cycle dependence of quiet-time magnetospheric currents and a model of their near-Earth magnetic fields, *Earth Planets and Space*, **62**(10), 843–848.

Lühr, H., Maus, S., Rother, M., & Cooke, D., 2002. First in-situ observation of night-time F region currents with the CHAMP satellite, *Geophysical Research Letters*, **29**(10), 4.

Lühr, H., Rother, M., Maus, S., Mai, W., & Cooke, D., 2003. The diamagnetic effect of the equatorial Appleton anomaly: Its characteristics and impact on geomagnetic field modeling, *Geophysical Research Letters*, **30**(17).

Lühr, H., Aylward, A., Stolle, C., Spain, T., Förster, M., Ritter, P., & Aruhlia, A., 2008. Ionospheric current quantification and modelling for improved magnetic and electric field analyses for Swarm, Final Report SWIO-SST-FR(2); ESTEC Contract No. 20943/07/NL/JA, ESA.

Lühr, H., Korte, M., & Mandea, M., 2008. The Recent Geomagnetic Field and its Variations, in *Geomagnetic Field Variations*, edited by K.-H. Glassmeier, H. Soffel, & J. Negendank, Advances in Geophysical and Environmental Mechanics and Mathematics, pp. 25–63, Springer Berlin Heidelberg.

Malin, S. R. C., 1982. Sesquicentenary of Gauss's First Measurement of the Absolute Value of Magnetic Intensity, *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **306**(1492), pp. 5–8.

Mandea, M., 2006. Magnetic satellite missions: where have we been and where are we going?, *Comptes Rendus Geoscience*, **338**(14-15), 1002–1011.

Mandea, M. & Olsen, N., 2006. A new approach to directly determine the secular variation from magnetic satellite observations, *Geophysical Research Letters*, **33**(15).

Maus, S. & Lühr, H., 2005. Signature of the quiet-time magnetospheric magnetic field and its

electromagnetic induction in the rotating Earth, *Geophysical Journal International*, **162**(3), 755–763.

Maus, S. & Lühr, H., 2006. A gravity-driven electric current in the Earth's ionosphere identified in CHAMP satellite magnetic measurements, *Geophysical Research Letters*, **33**(2), 5.

Maus, S., Lühr, H., Balasis, G., Rother, M., & Mandea, M., 2005. Introducing POMME, the Potsdam Magnetic Model of the Earth, *Earth Observation with Champ: Results from Three Years Orbit*, pp. 293–298.

Maus, S., Lühr, H., & Purucker, M., 2006. Simulation of the high-degree lithospheric field recovery for the Swarm constellation of satellites, *Earth Planets and Space*, **58**(4), 397–407.

Maus, S., Rother, M., Stolle, C., Mai, W., Choi, S., Lühr, H., Cooke, D., & Roth, C., 2006. Third generation of the Potsdam Magnetic Model of the Earth (POMME), *Geochemistry Geophysics Geosystems*, **7**.

Maus, S., Manoj, C., Rauberg, J., Michaelis, I., & Lühr, H., 2010. NOAA/NGDC candidate models for the 11th generation International Geomagnetic Reference Field and the concurrent release of the 6th generation POMME magnetic model, *Earth Planets and Space*, **62**(10), 729.

Menke, W., 1989. *Geophysical Data Analysis: Discrete Inverse Theory*, International Geophysics, Elsevier Science.

Millward, G., Moffett, R., Quegan, S., & Fuller-Rowell, T., 1996. A Coupled Thermosphere-Ionosphere-Plasmasphere model (CTIP), in *Solar-Terrestrial Energy Program: Handbook of ionospheric models*, edited by R. Schunk, Utah State University.

Mitra, S. K., 1946. Geomagnetic control of region-F2 of the ionosphere, *Nature*, **158**(4019), 668–669.

Monahan, A., Fyfe, J., Ambaum, M., Stephenson, D., & North, G., 2009. Empirical Orthogonal Functions: The medium is the message, *Journal of Climate*, **22**(24), 6501–6514.

Neubert, T., Mandea, M., Hulot, G., von Frese, R., Primdahl, F., Jørgensen, J. L., Friis-Christensen, E., Stauning, P., Olsen, N., & Risbo, T., 2001. Ørsted satellite captures high-precision geomagnetic field data, *Eos, Transactions American Geophysical Union*, **82**(7), 81–88.

Niblett, E. R. & Honkura, Y., 1980. Time-dependence of electromagnetic transfer functions and their association with tectonic activity, *Geophysical Surveys*, **4**, 97–114.

North, G. R., Bell, T. L., Cahalan, R. F., & Moeng, F. J., 1982. Sampling Errors in the Estimation of Empirical Orthogonal Functions, *Monthly Weather Review*, **110**(7), 699–706.

Olsen, N., 1997. Ionospheric F-region currents at middle and low latitudes estimated from Magsat data, *Journal of Geophysical Research-Space Physics*, **102**(A3), 4563–4576.

Olsen, N., 1999. Induction studies with satellite data, *Surveys in Geophysics*, **20**(3), 309–340.

Olsen, N., 2002. A model of the geomagnetic field and its secular variation for epoch 2000 estimated from Ørsted data, *Geophysical Journal International*, **149**(2), 454–462.

Olsen, N. & Kotsiaros, S., 2011. Magnetic Satellite Missions and Data, in *Geomagnetic Observations and Models*, edited by M. Mandea & M. Korte, vol. 5 of **IAGA Special Sopron Book Series**, pp. 27–44, Springer Netherlands.

Olsen, N. & Mandea, M., 2007. Investigation of a secular variation impulse using satellite data: The 2003 geomagnetic jerk, *Earth and Planetary Science Letters*, **255**(1-2), 94–105.

Olsen, N. & Mandea, M., 2008. Rapidly changing flows in the Earth's core, *Nature Geoscience*, **1**(6), 390–394.

Olsen, N. & Stolle, C., 2012. Satellite geomagnetism, *Annual Review of Earth and Planetary Sciences*, **40**, 441–465.

Olsen, N., Tøffner-Clausen, L., Sabaka, T. J., Brauer, P., Merayo, J. M. G., Jorgensen, J. L., Leger, J. M., Nielsen, O. V., Primdahl, F., & Risbo, T., 2003. Calibration of the Ørsted vector magnetometer, *Earth Planets and Space*, **55**(1), 11–18.

Olsen, N., Vennerstrom, S., & Friis-Christensen, E., 2003. Monitoring magnetospheric contributions using ground-based and satellite magnetic data, *First Champ Mission Results for Gravity, Magnetic and Atmospheric Studies*, pp. 245–250.

Olsen, N., Haagmans, R., Sabaka, T. J., Kuvshinov, A., Maus, S., Purucker, M. E., Rother, M., Lesur, V., & Mandea, M., 2006. The Swarm End-to-End mission simulator study: A demonstration of separating the various contributions to Earth's magnetic field using synthetic data, *Earth Planets and Space*, **58**(4), 359–370.

Olsen, N., Lühr, H., Sabaka, T. J., Mandea, M., Rother, M., Tofner-Clausen, L., & Choi, S., 2006. CHAOS – a model of the Earth's magnetic field derived from CHAMP, Ørsted, and SAC-C magnetic satellite data, *Geophysical Journal International*, **166**(1), 67–75.

Olsen, N., Sabaka, T., Gaya-Pique, L., Kuvshinov, A., & Tøffner-Clausen, L., 2007. Study of an Improved Comprehensive Magnetic Field Inversion Analysis for Swarm: Final Report, Tech. Rep. DNSC Scientific Report 1/2007, Danish National Space Center, Copenhagen.

Olsen, N., Mandea, M., Sabaka, T. J., & Tøffner-Clausen, L., 2009. CHAOS-2 – a geomagnetic field model derived from one decade of continuous satellite data, *Geophysical Journal International*, **179**(3), 1477–1487.

Olsen, N., Glassmeier, K. H., & Jia, X., 2010. Separation of the Magnetic Field into External and Internal Parts, *Space Science Reviews*, **152**(1-4), 135–157.

Olsen, N., Hulot, G., & Sabaka, T. J., 2010. Measuring the Earth's Magnetic Field from Space: Concepts of Past, Present and Future Missions, *Space Science Reviews*, **155**(1-4), 65–93.

Olsen, N., Mandea, M., Sabaka, T. J., & Toffner-Clausen, L., 2010. The CHAOS-3 geomagnetic field model and candidates for the 11th generation IGRF, *Earth Planets and Space*, **62**(10), 719–727.

Olsen, N., Kuvshinov, A., Macmillan, S., Püthe, C., Sabaka, T., Stolle, C., & Velimsky, J., 2011. Swarm Level 2 Processing System; Scientific Report on CAT-3 Activities, Final Report SW-RP-DTU-GS-0008; Rev: 1A, ESA.

Olson, P., Christensen, U., & Glatzmaier, G. A., 1999. Numerical modeling of the geodynamo: Mechanisms of field generation and equilibration, *Journal of Geophysical Research: Solid Earth (1978–2012)*, **104**(B5), 10383–10404.

Parker, R. L. & Shure, L., 1982. Efficient modeling of the Earth's magnetic field with harmonic splines, *Geophysical Research Letters*, **9**(8).

Perrone, L. & De Franceschi, G., 1998. Solar, ionospheric and geomagnetic indices, *Annali di Geofisica*, **41**(5-6), 843–855.

Pfaff, R. F., 2012. The Near-Earth Plasma Environment, *Space Science Reviews*, **168**(1-4), 23–112.

Preisendorfer, R. W. & Mobley, C. D., 1988. *Principal component analysis in meteorology and oceanography*, Elsevier; Elsevier Science Pub. Co., Amsterdam.

Press, W., 1992. *Numerical Recipes in Fortran 77: The Art of Scientific Computing*, Fortran Numerical Recipes, University Press.

Reigber, C., Lühr, H., & Schwintzer, P., 2002. *CHAMP mission status*, vol. 30 of **Advances in Space Research**, pp. 129–134, Pergamon-Elsevier Science Ltd, Oxford.

Richman, M. B., 1986. Rotation of principal components, *Journal of Climatology*, **6**(3), 293–335.

Richmond, A. D., 1995. Ionospheric electrodynamics using magnetic apex coordinates, *Journal of Geomagnetism and Geoelectricity*, **47**(2), 191–212.

Rishbeth, H., 1971. F-layer dynamo, *Planetary and Space Science*, **19**(2), 263–267.

Rishbeth, H., 1988. Basic physics of the ionosphere - a tutorial review, *Journal of the Institution of Electronic and Radio Engineers*, **58**(6), S207–S223.

Rishbeth, H., 1997. The ionospheric E-layer and F-layer dynamos – a tutorial review, *Journal of Atmospheric and Solar-Terrestrial Physics*, **59**(15), 1873–1880.

Ritter, P. & Lühr, H., 2006. Curl-B technique applied to Swarm constellation for determining field-aligned currents, *Earth Planets and Space*, **58**(4), 463–476.

Russell, C., 2000. The solar wind interaction with the Earth's magnetosphere: a tutorial, *Plasma Science, IEEE Transactions on*, **28**(6), 1818 –1830.

Russell, C. T. & Jian, L., 2008. Flows and obstacles in the solar wind, *Advances in Space Research*, **41**(8), 1177–1187.

Sabaka, T. & Baldwin, R., 1993. Modeling the Sq magnetic field from POGO and Magsat satellite and contemporaneous hourly observatory data: phase I, *Contract Report HSTX/G&G9302*.

Sabaka, T. J., Olsen, N., & Langel, R. A., 2002. A comprehensive model of the quiet-time, near-Earth magnetic field: phase 3, *Geophysical Journal International*, **151**(1), 32–68.

Sabaka, T. J., Olsen, N., & Purucker, M. E., 2004. Extending comprehensive models of the Earth's magnetic field with Ørsted and CHAMP data, *Geophysical Journal International*, **159**(2), 521–547.

Schunk, R. W. & Nagy, A., 2009. *Ionospheres : physics, plasma physics, and chemistry*, Cambridge University Press, Cambridge.

Selva, D. & Krejci, D., 2012. A survey and assessment of the capabilities of Cubesats for Earth observation, *Acta Astronautica*, **74**, 50–68.

Shore, R. M., Whaler, K. A., Macmillan, S., Beggan, C., Olsen, N., Spain, T., & Aruliah, A., 2013. Ionospheric midlatitude electric current density inferred from multiple magnetic satellites, *Journal of Geophysical Research: Space Physics*, **118**(9), 5813–5829.

Stauning, P. & Primdahl, F., 2000. First detection of global dawn-dusk ionospheric current intensities using Ampere's integral law on Ørsted orbits, *Geophysical Research Letters*, **27**(20), 3273–3276.

Stolle, C., Lühr, H., Rother, M., & Balasis, G., 2006. Magnetic signatures of equatorial spread F as observed by the CHAMP satellite, *Journal of Geophysical Research-Space Physics*, **111**(A2).

Stolle, C., Manoj, C., Lühr, H., Maus, S., & Alken, P., 2008. Estimating the daytime equatorial ionization anomaly strength from electric field proxies, *Journal of Geophysical Research-Space Physics*, **113**(A9), 13.

Suzuki, A. & Fukushima, N., 1982. Sunward or anti-sunward electric current in space below the MAGSAT level, *Geophysical Research Letters*, **9**(4), 345–347.

Tarduno, J. A., Cottrell, R. D., Watkeys, M. K., & Bauch, D., 2007. Geomagnetic field strength 3.2 billion years ago recorded by single silicate crystals, *Nature*, **446**(7136), 657–660.

Tarpley, J. D., 1970. Ionospheric wind dynamo .2. Solar tides, *Planetary and Space Science*, **18**(7), 1091–1103.

Vanhamaki, H. & Amm, O., 2011. Analysis of ionospheric electrodynamic parameters on mesoscales - a review of selected techniques using data from ground-based observation networks and satellites, *Annales Geophysicae*, **29**(3), 467–491.

von Storch, H. & Navarra, A., 1999. *Analysis of climate variability : applications of statistical techniques : Proceedings of an autumn school organized by the commission of the European Community on Elba from October 30 to November 6, 1993*, Springer, Berlin.

von Storch, H. & Zwiers, F., 2002. *Statistical Analysis in Climate Research*, Cambridge University Press.

Walker, M. & Jackson, A., 2000. Robust modelling of the Earth's magnetic field, *Geophysical Journal International*, **143**, 799–808.

Weisstein, E., 2003. *CRC Concise Encyclopedia of Mathematics, Second Edition*, Taylor & Francis Group.

Wilks, D. S., 2005. *Statistical Methods in the Atmospheric Sciences, Volume 91, Second Edition (International Geophysics)*, Academic Press, 2nd edn.

Wood, B. J., Walter, M. J., & Wade, J., 2006. Accretion of the Earth and segregation of its core, *Nature*, **441**(7095), 825–833.

Xu, W. Y., 2008. Uncertainty in magnetic activity indices, *Science in China Series E-Technological Sciences*, **51**(10), 1659–1664.

Xu, W. Y. & Kamide, Y., 2004. Decomposition of daily geomagnetic variations by using method of natural orthogonal component, *Journal of Geophysical Research-Space Physics*, **109**(A5), 11.